DOCUMENT RESUME

ED 376 824 IR 055 275

AUTHOR Moen, William E.; McClure, Charles R.
TITLE The Government Information Locator Service (GILS): Expanding Research and Development of the ANSI/NISO Z39.50 Information Retrieval Standard. Final Report.
INSTITUTION Geological Survey (Dept. of Interior), Reston, Va.; Syracuse Univ., N.Y. School of Information Studies.
PUB DATE 7 Sep 94
NOTE 259p.; Funded by the Interagency Working Group on Data Management for Global Change. For related reports, see ED 326 247 and ED 349 031.
AVAILABLE FROM National Information Standards Organization (NISO), NISO Press Fulfillment, P.O. Box 338, Oxon Hills, MD 20750-0338.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC11 Plus Postage.
DESCRIPTORS *Access to Information; *Computer Networks; Computer System Design; *Federal Government; *Government Publications; Information Dissemination; Information Management; Information Processing; Information Retrieval; *Public Agencies; Public Policy; *Standards
IDENTIFIERS *Government Information Locator Service; Internet; Inventory Systems; National Information Policy

ABSTRACT

The Government Information Locator Service (GILS) is a response to the need of users to be able to identify, locate, and access or acquire publicly available federal information resources. GILS uses ANSI/NISO Z39.50, the American National Standard for information retrieval, and other relevant standards to support the deployment of agency-based, network accessible locators. The third in a series of projects conducted by study teams at Syracuse University on federal locator systems, the current project produced: (1) an application profile for the use of Z39.50 in GILS; (2) a background document describing design decisions and assumptions of the project team and used to inform stakeholders and build consensus; (3) three reports describing technical and policy issues of concern to the GILS initiative; and (4) this final report, which details the project's activities, identifies remaining issues for future GILS activities, and recommends a series of next steps. The completion of this project has laid the technical groundwork for GILS implementations and identified two important next steps to move the GILS initiative forward: establishing an interoperability testbed for demonstrating interoperable GILS products; and continuing to stimulate the market for GILS products and services and encouraging federal agencies to implement GILS locators. The project also developed a list of recommendations that address both short-term requirements of GILS and its long-term viability. It is noted that, although the technical groundwork has been laid and a government-wide policy framework is being developed, future emphasis on GILS development will need to shift to individual agencies. Additional information related to the project is provided in 17 attachments. (BBM)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 376 824

The Government Information Locator Service (GILS): Expanding Research and Development on the ANSI/NISO Z39.50 Information Retrieval Standard

Final Report

of the
Cooperative Research Study
between the
School of Information Studies, Syracuse University
and
The United States Geological Survey
funded by
The Interagency Working Group on Data Management for Global Change

William E. Moen<wemoen@mailbox.syr.edu>
Charles R. McClure <cmcclure@suvm.acs.syr.edu>
Principal Investigators

School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244-4100
Telephone: (315) 443-2911
Fax: (315) 443-5806

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
William E. Moen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

September 7, 1994

IR055075

BEST COPY AVAILABLE

Copies of this report may be purchased from the National Information Standards Organization (NISO). To order this report, contact:

NISO Press Fulfillment
P.O. Box 338
Oxon Hills, MD 20750-0338
USA

Telephone US and Canada: 800-282-NISO (6476)
Outside the US and Canada: 301-567-9522; Fax 301-567-9553

TABLE OF CONTENTS

Acknowledgements ..... *i*

Executive Summary ..... *iii*

1. Introduction 1

2. ANSI/NISO Z39.50: A Standard for Information Retrieval 3

3. Articulating a Vision: The Design Document for GILS 3

4. Developing a Vision: Previous Research on Federal Information Locators 5

5. The Current Research Project 6

   5.1. Goals and Objectives 7
   5.2. Description of Activities 7
      5.2.1. Initiation and Coordination 7
      5.2.2. Developing the Profile 8
      5.2.3. Consensus Building 8
      5.2.4. Other Project Activities 9
   5.3. Products Developed 10

6. Development of the GILS Profile 10

   6.1. What is a Profile 11
   6.2. The GILS Profile 11
   6.3. Lessons Learned from Developing the Profile 12
   6.4. Impact on Product Development 12
   6.5. Open System Environment Implementors' Workshop 13
   6.6. The GILS Profile as a Federal Information Processing Standard 13

7. Evaluation of Project 14

   7.1. Consensus Among Stakeholders 14
   7.2. Impact on Z39.50 Standards Development and USMARC Specifications 15
   7.3. Opportunity for Generating Awareness and Stimulating Interest 16

8. Issues Remaining for Future GILS Activities 16

   8.1. Implementing GILS 17
   8.2. Bibliographic Control in the Networked Environment 18
   8.3. Accommodating the Installed Base of Technology 19
   8.4. Extensibility of GILS 20
   8.5. Federal Information Policy Issues 21
      8.5.1. OMB's Roles and Responsibility 21
      8.5.2. GILS and IRM Roles and Responsibility 22
      8.5.3. GILS and Federal Library Community Roles and Responsibility 22
      8.5.4. GILS and Archives and Record Managers Community Roles and Responsibility 23
   8.6. Policy and Technical Standards 23
   8.7. Evaluating the GILS Profiling Process and Products 24

9. Next Steps to Realize the Promise of GILS 25

   9.1. A GILS Interoperability Testbed 25
   9.2. Stimulating the Market 26

10. Final Recommendations 28

11. Moving Forward 31

Notes 32

References 33

Appendix A Project Team Members 36

Attachments

Attachment A Expanding Research and Development on the NISO Z39.50 Search and Retrieval Standard, Project Abstract (October 1, 1993)

Attachment B Application Profile for the Government Information Locator Service (GILS) (May 7, 1994)

Attachment C Using Z39.50 in an Application of the Government Information Locator Service (GILS): A Background Paper (May 7, 1994)

Attachment D Using the Z39.50 Information Retrieval Protocol in the Internet Environment *(Lynch)*

Attachment E The Government Information Locator Service (GILS): Report to the Information Infrastructure Task Force (May 2, 1994)

Attachment F Project Workstatement

Attachment G Meetings of GILS Project Team

Attachment H Stakeholder Contact List

Attachment I Critical Review of WAIS as an Application Tool for GILS (*FS Consulting*)

Attachment J Interoperability and Conformance Issues in the Development and Implementation of the Government Information Locator Service (GILS) (*Preston & Lynch*)

Attachment K Requirements for Accommodating Information Systems Information and Records Management Needs within the Proposal for a Government Information Locator Service (GILS) and its Z39.50 Application (*Bearman*)

Attachment L Working Implementation Agreements for Open Systems Environment: Part 31 — Application Profile for the Government Information Locator Service (GILS) — Library Applications Special Interest Group

Attachment M Proposed Federal Information Processing Standard (FIPS) for Application Profile for the Government Information Locator Service (GILS): Request for Comment

Attachment N Response to Stakeholder on Suitability of Z39.50 for GILS

Attachment O USMARC Proposal 94-9: Changes to the USMARC Bibliographic Format to Accommodate Online Systems and Services

Attachment P Building a Policy for Information Technology Standards (*Moen*)

Attachment Q GILS Forum, An Electronic Discussion Group

LIST OF FIGURES

Figure 1. GILS Design and Functional Characteristics 4

Figure 2. Key Areas of Consensus for Developing a GIILS 5

Figure 3. Next Steps for GILS: Recommendations 28

Figure 4. Key Factors Affecting GILS Implementation 31

Acknowledgements

This research project could not have been completed without the assistance and participation of a number of individuals and organizations. We especially acknowledge the support for this project from the Interagency Working Group on Data Management for Global Change. The project profited from the continuing interest and constructive suggestions and ideas offered by Eliot Christian, United States Geological Survey, who served as project liaison.

The GILS Profile project team deserves credit for the successful completion of this project. Team members devoted time from their already full schedules to hammer out the details of the GILS Profile. The team members included: Kevin Gamiel, Clearinghouse for Network Information Discovery and Retrieval; Ralph LeVan, OCLC; Denis Lynch, ESL, Inc.; Margaret St. Pierre, WAIS, Inc.; Madeleine Stovel, Research Libraries Group, Inc.; Eliot Christian, United States Geological Survey; Tim Gauslin, United States Geological Survey; Sue Ruddle, Defense Technical Information Center; Yesha Yelena, National Institute of Standards and Technology; Ray Denenberg, Z39.50 Maintenance Agency, Library of Congress.

Project consultants prepared important background papers that identified issues and solutions related to the GILS Profile and broader policy and technical considerations. The consultants were David Bearman, Cecilia Preston, and Francois Schiettecatte. Special appreciation and thanks go to Clifford Lynch who monitored and reviewed project activities and products as well as collaborating on one of the background papers.

A number of organizations made important contributions to the project. The Coalition for Networked Information provided opportunities at their meetings for discussion of project activities. The Coalition, specifically through the efforts of Paul Peters and Craig Summerhill, is also hosting a newly established public electronic discussion group devoted to the GILS initiative. The Network Development and MARC Standards Office at the Library of Congress provided meeting space for the project team. In addition, Sally McCallum and Rebecca Guenther of that office supplied expertise in the mapping of GILS data elements to USMARC. The National Information Standards Organization and its executive director, Patricia Harris, supported the project by distributing GILS papers to a meeting of the Automation Vendor Information Advisory Committee as well as taking on the task of distributing this final report through the NISO Press.

Members of the Z39.50 Implementors Group (ZIG) and the subscribers to the USMARC discussion list provided important expertise on specific technical issues of the GILS Profile. Many ZIG members were willing to critique the GILS Profile work, and their constructive comments were essential to producing an implementable profile.

At the School of Information Studies, Kathleen Flynn was a graduate research assistant with many support responsibilities. She also provided editorial assistance in reviewing the final draft of this report. Beth Mahoney provided word processing and graphics support. Lori Brownell competently handled the many administrative activities related to the project. Rebecca Freeland helped coordinate travel, meetings, and timely reimbursements. The Principal Investigators also recognize the institutional support provided by Syracuse University that contributed to the completion of the study. We also want to thank the ERIC Clearinghouse and its AskERIC staff, particularly Michael Eisenberg and David Lankes for providing technology support for the electronic communication needs of the project team.

Finally, we would like to thank the many people with whom we talked about this project and the GILS Profile. The consensus building aspect of this project brought us in contact with people who are concerned about and interested in seeing a new era in access and dissemination of government information.

Although the authors gratefully acknowledge the assistance and contributions of all the people mentioned above, the opinions and recommendations expressed in this report are those of the Principal Investigators. These opinions and recommendations included here may not necessarily represent the views of the Interagency Working Group on Data Management for Global Change nor all members of the GILS project team.

William E. Moen
Charles R. McClure
September 1994

Executive Summary

The Government Information Locator Service (GILS) is a response to the need of users to be able to identify, locate, and access or acquire publicly available Federal information resources. GILS uses ANSI/NISO Z39.50, the American National Standard for information retrieval and other relevant standards to support the deployment of agency-based, network-accessible locators. This cooperative research project between the United States Geological Survey and Syracuse University developed an application profile for ANSI/NISO Z39.50 for use in GILS. The current project can be considered the third in a series of projects conducted by study teams at Syracuse University on Federal locator systems (see McClure, et al., 1990; McClure, Ryan & Moen, 1992).

This project began in September 1993 and was completed in May 1994. Objectives of this project included:

- Expand research and development on the American National Standard for information searching and retrieval (Z39.50) for its application in facilitating public access to Federal information resources and speeding the development of interoperable systems. This involved working within the voluntary standards system to investigate how Z39.50 and other standard could be used in GILS.

- Build consensus of major stakeholders on the manner in which Z39.50 can be applied in GILS implementations. This involved ongoing and wide-ranging information dissemination concerning the work and direction of the project, and included targeted mailings and presentations to stakeholders to build consensus on the specifications of a system architecture for GILS and the specifications of Z39.50 and other standards for use in GILS.

- Develop an application profile for networked-based GILS implementations that references Z39.50 and other standards for use in the Internet environment. This was a primary activity of the project and involved the work of a project team coordinated by the Principal Investigators to precisely identify and detail the specifications for Z39.50, data content standards, USMARC, and other standards. These specifications are included in the profile document, "Application Profile for the Government Information Locator Service (GILS)." The GILS Profile is now being processed by the National Institute of Standards

and Technology as a Federal Information Processing Standard (FIPS), and it has been approved by the Open Systems Environment Implementors Workshop as "Working Implementation Agreements for Open Systems Environment: Part 31 — Application Profile for the Government Information Locator Service (GILS) — Library Applications Special Interest Group."

- Support and encourage test implementations of the profile by interested parties to provide evaluations of the profile and for interoperability testing. This has included the identification of next steps for GILS implementation including an interoperability testbed and stimulating a market for GILS products and services.

The current project produced (1) an application profile for the use of Z39.50 in GILS; (2) a background document describing design decisions and assumptions of the project team and used to inform stakeholders and build consensus; (3) three reports describing technical and policy issues of concern to the GILS initiative such as interoperability testing, extensibility of GILS, and the accommodation of the installed based of technology; and (4) this final report that details the project's activities, identifies remaining issues for future GILS activities, and recommends a series of next steps. This report contains a series of attachments that include all relevant documents produced by the project or are otherwise pertinent to the GILS initiative. By including these attachments, this complete document serves as a comprehensive source of GILS-related information important to understanding the context and results of this research project.

The successful completion of this project has laid the technical groundwork for GILS implementations. This project identified two important next steps to move the GILS initiative forward: establishing an interoperability testbed for demonstrating interoperable GILS products; and continuing to stimulate the market for GILS products and services and encouraging Federal agencies to implement GILS locators. The project also developed a list of recommendations that address both short-term requirements of GILS (e.g., developing guidelines for GILS record creation) and the long-term viability of GILS (e.g., strategically placing GILS as part of major Federal initiatives such as the National Information Infrastructure).

Although the technical groundwork has been laid (i.e., the GILS Profile) and a government-wide policy framework is being developed (e.g., the Office of Management and Budget will soon release an OMB Bulletin on GILS), future emphasis on GILS development will need to shift to individual agencies. The Principal Investigators conclude that there are social, cultural, and organizational factors in Federal agencies that will affect the success of GILS development. Cultural and organizational changes need to be encouraged to realize the promise of GILS and its utility in Federal information resources management and as a vital mechanism providing access to government information.

The Government Information Locator Service (GILS): Expanding Research and Development on the ANSI/NISO Z39.50 Information Retrieval Standard

1. Introduction

The emerging National Information Infrastructure (NII) provides new opportunities for Federal agencies to disseminate publicly available government information and for the public to have the means for easy access to that information, especially when the information is held in electronic formats. Political leaders and policymakers recognize that access to government information is essential and have exhorted Federal agencies to improve such access. The recent Clinton Administration technology policy document, "Technology for America's Economic Growth: A New Direction to Build Economic Strength" (Clinton & Gore, 1993, p. 17) states:

> Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it.

As reflected in this statement, a major barrier to effective citizen access to public information is the lack of directories and other finding tools to identify and locate information resources that Federal agencies create, house, and disseminate.

The Government Information Locator Service (GILS) is a response to the needs of users to be able to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources. GILS uses ANSI/NISO Z39.50, the American National Standard for information retrieval (National Information Standards Organization, 1992), and other relevant standards to support the deployment of agency-based, network-accessible locators. These locators provide users with descriptive, location, and access information for a wide range of Federal government information resources. Z39.50 defines a standard way for two computers to communicate for the purpose of information retrieval and facilitates the use of large information databases by standardizing the procedures and features for searching and retrieving information (for a general overview of Z39.50, see, The ANSI/NISO Z39.50 Protocol: Information Retrieval in the Information Infrastructure [Moen, 1994]).

To advance the development of GILS, the United States Geological Survey (USGS) entered into a cooperative agreement with the Syracuse University to coordinate a research project focused on the use of open systems standards to improve the utility of information searching and retrieval via computer communications networks (Attachment A is a project abstract). This document is a report on that research project. The project began in September 1993 and was completed in May 1994. The current research builds upon a previous study, Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators (McClure, Ryan & Moen, 1992; McClure, Moen & Ryan, 1992). The Government Information Locator Service (GILS) (Christian, 1994) served as the defining vision for this project by describing and defining what GILS is, its objectives, and service requirements.

The project had as its objectives to:

- Expand research and development on the American National Standard for information searching and retrieval (Z39.50) for its application in facilitating public access to Federal information resources and speeding the development of interoperable systems
- Build consensus of major stakeholders on the manner in which Z39.50 can be applied in GILS implementations
- Develop an application profile for networked-based GILS implementations that references Z39.50 and other standards for use in the Internet environment
- Support and encourage test implementations of the profile by interested parties to provide evaluations of the profile and for interoperability testing.

A primary focus of the project was the development of an application profile for using Z39.50 in GILS. Constructing a standards-based GILS, based on a widely accepted application profile, will increase the likelihood of interoperability and interworking among the agency implementations. Further, these implementations can provide linkage to the installed based of user-oriented, information-access tools available through public domain software on the Internet, library-based information services, and other networked-based information providers. Equally important, the standards-based GILS will ensure wider access to Federal information resources.

This research project broke new ground for the Z39.50 community of developers and implementors. It resulted in first fully specified application profile for Z39.50 (Attachment B is the GILS application profile). In addition, it raised new Z39.50 implementation and standards-related issues such as the use of schemas, tag sets, and Uniform Resource Identifiers (URIs).[1] In some cases the project provided potential solutions for these issues, while in other cases these issues have become part of the working agenda of the Z39.50 Implementors Group (ZIG), a Z39.50 implementors and users forum that is responsible for enhancing and extending the standard.

The project also served to increase awareness of the GILS initiative through active dissemination of project documents, presentations at meetings, and other publicity efforts. These activities also served to build consensus on the manner in which Z39.50 would be used in the GILS (for example, Attachment C is a background document on the development of the GILS Profile used during the project to inform stakeholders and build consensus). Finally, the project has attempted to stimulate the development of off-the-shelf products that are compliant with the GILS Profile as well as stimulating a market for such products.

Although there remain important questions and actions related to the use of Z39.50, the implementation of GILS, and the organization and creation of locator records for GILS, this project has successfully provided the foundation for a series of next steps. The purpose of this final report is to describe the activities of the research project, to propose a series of next steps, and to discuss the implementation and policy issues that remain.

This report contains a number of attachments that were produced in the course of the study or are pertinent to the GILS initiative. By including these attachments, this complete document serves as a comprehensive source of GILS-related information relevant to understanding the context and results of this research project.

2. ANSI/NISO Z39.50: A Standard for Information Retrieval

The information retrieval protocol, Z39.50, provides a common language for clients to select and retrieve records from a range of servers. The purpose of Z39.50 is to allow one computer operating in a client mode to perform information retrieval queries against another computer acting as an information server and to provide for the transfer of records or other information from the server to the client. Z39.50 does not prescribe how a particular system will execute the searching and retrieval on databases nor does it prescribe user interface requirements.

Z39.50 is an applications-layer protocol originally modelled within the Open Systems Interconnection (OSI) Basic Reference Model. The OSI Basic Reference Model (ISO 7498: 1984 Open Systems Interconnection--Basic Reference Model) was developed at the international level by the International Organization for Standardization (ISO). Applications-layer protocols support the communications requirements of and interact directly with computer programs that reside on clients and servers and perform specific operations.

The National Information Standards Organization (NISO), an American National Standards Institute (ANSI) accredited standards developer that serves the library, information, and publishing communities, developed Z39.50. This standard was first approved as an American National Standard in 1988. NISO balloted and approved a 1992 revision of the standard (Z39.50-1992, also referred to as Version 2). Since 1991, the ZIG has been preparing a new version of the standard, Z39.50-199x (sometimes referred to as Version 3). NISO began the balloting process on this new version of Z39.50 on September 1, 1994.[2]

Z39.50 is a compatible superset of two International Standards for information retrieval: ISO 10162, Search and Retrieve Application Service Definition and ISO 10163-1, Search and Retrieve Protocol Specification. In early 1994, international standards developers made a crucial decision to begin the process of converging the international standards with U.S. Z39.50 work. No longer will there be different national and international standards that must be harmonized. Rather, it is the intention of national and international standards developers to use the version of Z39.50 now being balloted as a basis for both the American and International Standards. A reflection of this intention has been the participation of international Z39.50 users at the April 1994 meeting of the ZIG; plans are underway to hold the Spring 1995 ZIG meeting in Europe.

Although modelled as an OSI applications-layer protocol, Z39.50 is currently used by implementors in the Internet environment. The success of a Z39.50 interoperability testbed in 1992 showed that the transport service of the Internet's Transmission Control Protocol (TCP) can successfully support the Z39.50 protocol. Lynch (1994a) describes how Z39.50 can be implemented over TCP. Attachment D contains a draft Internet Engineering Task Force (IETF) Request for Comment (RFC) by Lynch (1994b), "Using the Z39.50 Information Retrieval Protocol in the Internet Environment."

Initial Z39.50 applications supported information retrieval of bibliographic data, but a growing number of implementations are expanding the range of Z39.50 applications. In addition, commercial, off-the-shelf Z39.50 products are becoming increasingly available (for a list of vendors providing Z39.50 products, see Moen, 1994). The ZIG, a group of active Z39.50 implementors, continues to enhance and refine the standard based on the requirements of implementors and other users. Z39.50 is stable, implementable, and is proving to be an important tool for information retrieval in the emerging information infrastructure. As a technical component of GILS, Z39.50 will assist in moving GILS from vision to concrete implementations.

3. Articulating a Vision: The Design Document for GILS

The GILS initiative arises from the co-incidence of several forces including:

- The Clinton Administration's Strategic Technology policy statement (see Clinton & Gore, 1993)
- The Administration's commitment to the development of a National Information Infrastructure (see Information Infrastructure Task Force, 1993)
- A strengthened Federal policy on information resources management (IRM) (see Office of Management and Budget, 1993a, 1994)
- The increasing role of networking technology available to increased numbers of users
- Federal agencies becoming connected to and experienced with the Internet.

Earlier research (see Section 4) provided a basis for a vision of a government-wide locator, and Christian (1994) articulated more explicitly a design for GILS.

Government Information Locator Service (GILS): Report to the Information Infrastructure Task Force (Christian, 1994) describes the vision and function of GILS and outlines its objectives and service requirements (a copy of this document is included in Attachment E; throughout this report this item is referred to as the GILS document). Drafts of the GILS document circulated to Federal agencies, interested and potential stakeholders, and the public from Fall 1993 through Spring 1994. While the defining vision of GILS in that document remained relatively stable, the iterations of the drafts developed details on implementation and aided in the evolution of an understanding how GILS might be implemented. The implications for this research project of the evolving understanding of GILS are discussed below.

GILS is a response to the need for users to be able to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources. It is a decentralized collection of locators and associated information services that includes information and technology components as well as policy, legal and regulatory mandates, and people. GILS is intended to help the public locate and access public information throughout the U.S. government. Based on the GILS document, GILS implementations are to exhibit design and functional characteristics. Figure 1 summarizes the characteristics (based on the GILS document) that were particularly relevant for the current research project. The GILS document addresses additional design, operational, and other criteria (e.g., intellectual property safeguards, privacy concerns).

To move from a vision of the GILS to actual implementations, the design and high-level functional requirements outlined in the GILS document — especially those related to the use of Z39.50 — needed to be specified in more detail. The exact manner in which GILS would use Z39.50 and other related or emerging standards needed explicit definition to increase the likelihood of interoperability and to provide users the capability to search across the vast information space of Federal information resources.

As context for this project, the next section summarizes previous research undertaken by Syracuse University related to Federal information locators. The research efforts over the past six years suggest that a user-based approach to designing a locator system, policy analysis and policy advocacy, and an awareness of technology trends and information technology standards can be particularly effective to connect the

Figure 1. GILS Design and Functional Characteristics

- Comprehensive in its coverage of Federal information resources
- User friendly
- Answers specific questions
- Allows scanning of a wide range of government information
- Responds to needs and abilities of naive users as well as sophisticated researchers
- Provides service directly to the public
- Does not undermine the diversity of existing information sources
- Can be used either directly or through intermediaries
- Provides information regarding request and delivery of referenced information resources
- Equipment and software requirements, cost, and technical complexity must be minimized as barriers
- Uses network technology to connect distributed servers that are agency-based
- Conforms to national and international standards for information and data processing
- Supports seamless access not only among locators but directly to the referenced information resources
- Defines a subset of all GILS components and refers to this as the GILS Core; the GILS Core comprises those locator records maintained by the U.S. Federal government, all of which comply with the defined GILS Core Element standards; GILS Core Elements define the content of a finite number of data elements used in individual locator records to describe information resources
- Must be accessible on interconnected electronic network facilities and must support the currently approved ANSI/NISO Z39.50 standard for information retrieval
- Must conform to the GILS Profile to provide full functionality to GILS direct users

needs of information users and providers in achieving broader Federal information policy goals.

4. Developing a Vision: Previous Research on Federal Information Locators

In 1990, Syracuse University researchers conducted the first of two studies related to improving the public's knowledge and access to Federal information resources through the use of Federal information inventories and locators. The General Services Administration/Regulatory Information Service Center (GSA) and the Office of Management and Budget, Office of Information and Regulatory Affairs (OMB-OIRA) funded the study (referred to as Phase I of the two-part research effort).

Federal Information Inventory/Locator Systems: From Burden to Benefit (McClure et al., 1990), the final report from that study, included a policy analysis of locator-related legislation and related policy instruments, assessments from a number of key Federal officials, and an analysis of public comments on how a locator of Federal information resources might be developed. One of the findings from the research was the need to add a government-wide, electronically-based finding tool to the traditional finding aids currently available. The study also concluded that the Federal Information Inventory Locator System (FILS) mandated in 44 U.S.C. 3507-3511 was ineffective and inadequately addressed access to and improved dissemination of government information. The study suggested that a new approach was needed as a means to identify, locate, and obtain government information. The approach recommended in the study was a government-wide information inventory/locator system (GIILS) that would address objectives such as access and dissemination, which were quite different than the objectives of the FILS.[3]

The study also revealed that a range of stakeholder groups expressed widespread interest in the development of a GIILS. While there was not consensus on the technical design details of a GIILS, the stakeholders agreed that a GIILS should be designed in light of the basic principles listed in Figure 2 (McClure et al., 1990).

Several of the areas identified in the study as needing further investigation were of interest to OMB and the National Archives and Records Administration (NARA) and included: 1) the identification of existing locator systems; 2) the creation of a machine-readable database housing descriptions of these locator systems; and 3) the identification and discussion of key issues related to the actual development of a government-wide locator system. GSA, NARA, and OMB/OIRA, jointly funded Phase II of the study to address these areas. Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators (McClure, Ryan & Moen, 1992) is the final report from Phase II.

The Phase II study began in May 1991 and concluded in August 1992. It had the following specific objectives:

- Identify existing and planned Federal agency locators
- Identify critical success factors in the design, development, and maintenance of a Federal agency locator

Figure 2. Key Areas of Consensus for Developing a GILS

- The Government should be responsible for GIILS development
- OMB should develop and enforce clear and consistent GIILS policy guidelines but should not be involved in the actual operation of a GIILS
- The system must respond to user information needs
- The GIILS design and operation should be based on input from a range of stakeholders
- Standards for operations and performance be identified and maintained
- The agencies should be the locus of responsibility and control
- Agencies should have incentives and receive rewards for participating in GIILS
- Any GIILS should be integrated into agency information resources management (IRM) functions
- Congress must provide support for a GIILS
- Keep the GIILS simple and develop it incrementally
- GIILS should provide multiple products in a range of formats

- Identify and discuss key issues and policy recommendations related to the design and development of a government-wide locator system.

The study recommended that OMB develop a policy framework requiring agencies to design and maintain machine-readable locators that would be accessible over the Internet and that would meet certain requirements and standards.

The study developed a definition of a "locator," based in part on the types of locators identified and assessed in the study and on the developments of network technologies including the rapid growth of the Internet and tools available for use in the networked environment. Minimally, a locator should meet certain criteria (McClure, Ryan, & Moen, 1992, p. 2):

- The locator is a point of entry for locating government information, regardless of the format and content of that information
- The content of the locator is not the actual information resource or service itself; rather it is a description of that information
- The locator tells the user (1) what information is available on a particular topic, (2) where that information is located, and (3) how the user would access that information
- The locator is in machine-readable format
- The locator is publicly accessible and searchable—preferably through direct dial-up telephone lines and/or through the Internet/NREN.

This definition informs the current GILS initiative.

Acknowledging the decentralized nature of Federal information resources management (IRM), the study recommended a strategy that would build upon this decentralized management context by having the individual agencies be responsible for developing and implementing locators to their own information resources. Congressional mandate and Executive regulations (e.g., 44 U.S.C., Freedom of Information Act, OMB Circular A-130) already required agencies to establish and maintain a range of information inventories and locator systems. The locator system envisioned in Phase II would use computer and communications technologies to integrate these agency-based locators into a "virtual" government-wide information locator system.

The study identified client/server architecture and Z39.50 for networked information retrieval as key technology components for realizing a distributed, virtual locator system. By using Z39.50 when implementing their locators, agencies would lay the foundation for transparent navigation and access through the vast range of information housed on the individual agencies' locators.

Phase II did not detail the technical specifications for an agency-based, network-accessible government locator system. Instead, it painted a vision of what such a system might look like and some of the essential components. The study also addressed the importance of establishing a policy framework for the locator system. Although the technology had emerged or was emerging to support such a vision, technical solutions themselves needed to be guided by explicit and coordinating policy decisions.

The discussion in Section 3 presented a further articulation of the vision from the Phase II study. The design document (Christian, 1994) for GILS provided the basis for the research project that is the focus of this report, a project that has moved the vision of GILS to actual technological implementations using Z39.50.

5. The Current Research Project

To realize the vision of transparent network access to government information via a system of agency-based locators, there was a need for additional research on the development and implementation of Z39.50 for use in a locator application. The choice of Z39.50 as the appropriate national standard for use in GILS required additional specification of how GILS would work as well as which specific features of Z39.50 would support the functionality required in GILS.

Computer communications protocols are complex technical specifications that often contain a variety of options and choices from which to make selections when developing an implementation (McCallum, 1994). To increase the likelihood of achieving the desired interoperability of independently developed Z39.50 GILS components, implementors must agree on which options, choices, and features from Z39.50 that GILS implementations would support. A profile is a mechanism for accomplishing this; a profile is a set of implementation agreements that guide implementors in applying one or more standards in a specific and limited context.

Defining a GILS Profile would be an important contribution to Z39.50 use in GILS. In addition, some GILS requirements as described in the GILS document demanded Z39.50 implementations needing functionality not specifically addressed in the standard. Thus, there was a likelihood that GILS requirements related to Z39.50 would need to be addressed by the Z39.50 standards developers (i.e., the ZIG). Further, the GILS Profile would need to address concerns such a data content standards (e.g., the GILS Core Elements) that are beyond the scope of Z39.50 and conventional profiles. GILS Profile development, in effect, would redefine the currently accepted definition of what a profile addresses. The resulting GILS Profile is more of a "system" profile than a profile for a single standard.

Syracuse University and the United States Geological Survey (USGS) entered into a cooperative agreement funded by the Interagency Working Group on Data Management for Global Change to conduct this research and development. The study began in September 1993 and concluded in May 1994. The primary focus of the project concerned the development of an application profile for the GILS. In addition, the project stressed the importance of outreach to interested and potential stakeholders through direct communication and dissemination of ongoing research results. Another aspect of this research and development effort was to promote wider acceptance of Z39.50, specifically, its use in GILS. Moreover, the project intended to increase awareness of the importance of Z39.50 as applied to the search and retrieval of locator and other information.

The project successfully completed a stable draft of the GILS Profile, a document that fully specifies the use of Z39.50 in an application of the GILS. This section of the report describes in more detail the project, its activities, and its products.

5.1. Goals and Objectives

For the current study, the Principal Investigators coordinated the work of a group comprising technical experts (e.g., in Z39.50 implementations, information systems implementations, and information organization) and representatives of Federal agencies. Guided by an overriding goal to advance the development of GILS, the research project focused on the use of open systems standards to improve the utility of information searching and retrieval on digital networks. More specifically, the project had as its objectives to:

- Expand research and development on the American National Standard for information searching and retrieval (Z39.50) for its application in facilitating public access to Federal information resources and speeding the development of interoperable systems
- Build consensus of major stakeholders on the manner in which Z39.50 can be applied in GILS implementations
- Develop an application profile for networked-based GILS implementations that references Z39.50 and other standards for use in the Internet environment
- Support and encourage test implementations of the profile by interested parties to provide evaluations of the profile and for interoperability testing.

The development of the GILS Profile and building consensus on the manner in which Z39.50 would be used were the necessary first steps in achieving any subsequent objectives. The project team focused its primary attention on these two activities. Given the length of the project, the concluding activities have attempted to stimulate interest (both user and vendor) in implementations that conform to the GILS profile.

5.2. Description of Activities

The activities for the project can be grouped into the following categories:

- Initiation and Coordination
- Developing the Profile
- Consensus Building

The next sections briefly describe how the Principal Investigators carried out these activities.

5.2.1. Initiation and Coordination

To initiate the research and development effort, project leaders from Syracuse University and USGS met and discussed the general direction, potential participants, and activities for the project. They concluded that a project team strategy would be most effective in accomplishing the work and objectives of the project.

The project's focus on developing the GILS Profile required the involvement of experts in Z39.50 and information systems. These experts would nominally represent the communities of interested or potentially interested technology providers and users. Representatives from Federal agencies were also needed to provide nominal representation from the perspective of GILS providers (i.e., the agencies that would be creating the locator records, implementing GILS locators, etc.).

The project team included experts in Z39.50, information systems, organization and management of information resources, and representatives of Federal agencies. Appendix A lists the project team members. The project team, as nominal representatives of several communities of interest, also served as a basis for consensus building on GILS. Constraints of time and budget required a limited — but representative and interested — number of participants on the team. Syracuse project leaders developed a workstatement for the project team, and contracted with participants to perform the duties outlined in the workstatement (see Attachment F). These tasks reflected the overall objectives for the project.

The project team primarily carried out its responsibilities in a series of working meetings. Attachment G lists the meetings of the project team and other meetings where a majority of the project team were present and discussions on the project occurred.

Syracuse project leaders established a electronic discussion forum for the use of team members and other interested individuals. The electronic discussion group proved to be an essential component to the deliberations of the team. It was a closed discussion forum where ideas and criticisms could be exchanged in a constructive manner by those committed to seeing the project reach a successful conclusion. Project leaders and team members used this discussion group to circulate drafts of documents, gather comments for revisions, post summaries of the team meetings, etc. The collaborative work of the project was substantially enhanced through the use of electronic networks.

Throughout the project, Syracuse project leaders communicated with team members, coordinated the arrangements of meetings, made travel arrangements and accommodations for the meetings, and carried out other project management activities.

5.2.2. Developing the Profile

The development of the GILS Profile occurred over three GILS project team meetings. This activity became central to the entire research project, and because of the lessons to be learned from this activity, it receives a more complete description in a separate section of this report (Section 6).

5.2.3. Consensus Building

A basic assumption governing the execution of the project was that widespread acceptance of a GILS Profile and its utility in implementations would be facilitated by building consensus among interested and potential stakeholders. Team members were nominal representatives of a number of potentially interested communities such as information services providers, libraries, technology and software developers and providers, the Z39.50 implementors community, and Federal agencies. Project leaders expected team members to use their connections with these various communities to gain input into the process of developing the GILS Profile.

A number of outreach efforts occurred to reach a wider group of potentially interested individuals and communities. The following lists these activities:

- Presentations at ZIG meetings (October 1993, Ottawa, Canada; January 1994, Gainesville, FL; April 1994, Washington, DC) to inform and update this primary stakeholder community on the work of the project team. Requests were made for input to the work of the project team. At the January 1994 meeting, a focus group discussion with Z39.50 implementors elicited concerns and issues based the January draft of the background document produced by the project team, "Using Z39.50 in an Application of the Government Information Locator Service (GILS)" (Attachment C).

- Electronic announcements began in January 1994 to a number of Internet electronic discussion groups which might include interested and potential stakeholder communities. These included: CNI-ANNOUNCE (an announcement service of the Coalition for Networked Information); GOVDOCS-L (listserv primarily serving govern-

ment documents librarians); PACS-L (library and technology oriented listserv); USMARC (a listserv for discussions of the USMARC format); and Z3950IW (the listserv of the ZIG).[4] These included announcements of the availability of project documents and requests for comments on specific work areas of the project team.

- Direct mailing of January draft, "Using Z39.50 in an Application of the Government Information Locator Service (GILS)," to selected attendees at the Fall 1993 Meeting of the Coalition for Networked Information (November 19-20, 1993, Chantilly, VA). The mailing also included a solicitation for comment on the document. Project description, goals, and objectives distributed at that meeting.

- Distribution of GILS material at the February 1994 meeting of the Automation Vendor Information Advisory Committee (AVIAC), Los Angeles, CA.

- Presentation to the E-Media Conference (January 20, 1994, Washington, DC) for Federal agency representatives to inform, update, and request feedback on the work of the project team. "Using Z39.50 in an Application of the Government Information Locator Service (GILS)" distributed at the meeting.

- Briefing on GILS project at the Spring 1994 Meeting of the Coalition for Networked Information (April 5, 1994, Washington, DC) to inform, update, and request feedback from meeting attendees. Summary sheet on project with pointers to electronic versions of project documents distributed.

Contact with potential stakeholders and interested parties via mail, telephone, personal conversation, and electronic mail served to inform and provided an opportunity to request input and comments on the project. Attachment H contains a list of contacts made in the course of the project.

Since the research project focused on the use of Z39.50 in GILS and developing an application profile that would specify the use of Z39.50, much of the team's work was technically oriented. This is reflected in the documents produced by the project team. These technical documents are not easily accessible to the uninitiated in Z39.50, and this could account for a relatively low response rate to the requests for comments and input into the project team's work.

Most productive to the team's contact with potential stakeholder groups was the interaction with the Z39.50 implementors and the people involved in USMARC format development and use. Z39.50 implementors have substantially agreed with the approach and specifics of the team's profile effort. This group was also most able to address the technical nature of the documentation.

In the case of the USMARC community, USMARC experts were requested to address one particular area of concern to the project team (i.e., the mapping of GILS Core Elements to USMARC fields). Approximately 25 individuals reviewed the proposed mapping, and ten individuals from the USMARC community forwarded recommendations. The response rate from the USMARC community was high because of the interest and expertise of these people in dealing with these specific technical issues of the GILS Profile.

Consensus building on the overall GILS design (as presented in the GILS document) was not the responsibility of the project team and project leaders at Syracuse. Eliot Christian, among others, played an important role in creating awareness of the GILS design and building consensus on that design. The project team and project leaders, however, received questions on certain design and policy issues presented in the GILS document. These policy and design issues (e.g., what information resources would be described in locators, who would make sure agencies implemented GILS locators, etc.) were beyond the scope of project team responsibilities. Some policy issues, however, are of concern to the successful implementation of GILS (e.g., bibliographic control, GILS flexibility and extensibility to accommodate new uses). These issues are detailed in Section 8 below.

5.2.4. Other Project Activities

In addition to the documents produced by the project team, three other project activities produced written reports. These reports helped increase the overall understanding of the issues and implications of implementing the GILS. Contractors, in addition to the core project team members who were developing the profile, completed these activities.

A technical report prepared by FS Consulting, "Critical Review of WAIS as an Application Tool for GILS," (Attachment I) examined the use of Wide Area Information Server (WAIS) technology in GILS. WAIS

implementations were originally based on Z39.50-1988 with extensions defined by RFC 1625, WAIS over Z39.50-1988 (St. Pierre, 1994). A WAIS application profile now specifies how WAIS will be implemented using Z39.50-1992. A number of Federal agencies have deployed or are deploying WAIS technology as part of their efforts in providing access to government information, and it was important to understand the extent to which WAIS might effectively support GILS specifications.

A major objective of GILS is to provide for interoperability of separately developed components from various vendors and implementors using different platforms and architectures. Namely, GILS is to operate in an Open Systems Environment, which is a "computing environment that supports portable, scalable, and interoperable applications through standard services, interfaces, data formats, and protocols" (National Institute of Standards and Technology, 1994, p. E-1). A technical report by Cecilia Preston and Clifford Lynch, "Interoperability and Conformance Issues in the Development and Implementation of the Government Information Locator Service (GILS)," (Attachment J) examined the issues related to interoperability and conformance testing for GILS implementations.

GILS flexibility and its ability to accommodate new requirements (i.e., be extensible) is also an important consideration. GILS will be an evolving system, and the design and specifications for using Z39.50 had to respond to uncertain evolution. One way in which this could be tested was to ask for input from a potential user community about its needs which, according to at least one spokesperson, were not being fully addressed in the initial design of the GILS. A technical report prepared by David Bearman, "Requirements for Accommodating Information Systems Information and Records Management Needs within the Proposal for a Government Information Locator Service (GILS) and its Z39.50 Application," (Attachment K) examined the needs of the archives and records management communities to understand how GILS might be extended to address additional requirements. This report also provided a way to ensure that nothing in the specifications of the GILS Profile would preclude such extensions.

These three activities, along with the documents produced by the project team, provide a sense of the scope of issues, concerns, and implications of GILS deployment.

A final component of the project was the use of a outside reviewer to examine and comment on the process and products of the project team and project consultants. Clifford A. Lynch, a respected expert in Z39.50 standards development and implementation, information systems, and policy, critically reviewed and made suggestions throughout the project.

5.3. Products Developed

The primary product of the project was the profile, "Application Profile for the Government Information Locator Service (GILS)." This GILS Profile provides the complete specification for the use of Z39.50 in GILS. Another important product was a document, "Using Z39.50 in an Application of the Government Information Locator Service (GILS)," that discussed the development of the GILS Profile and detailed the assumptions made by project team members about GILS, the system architecture model, and the resulting choices and determinations about Z39.50 features that would be used by GILS. The purpose of the background document was to provide stakeholders with adequate information so that they could understand the reasons why the project team made certain design and other choices for Z39.50 and GILS.

These documents circulated among the project team members and were distributed, either electronically or in paper copy, to external stakeholders. Each document went through a number of revisions based on comments received. Attachment B and Attachment C include the final versions of these documents.

6. Development and Approval of the GILS Profile

This section describes the development of the GILS Profile. For the Z39.50 community, profiling is a new activity.[5] In the OSI environment, "profiles" developed as an auxiliary mechanism to assist implementors in using one or more standards in an application and because of the complexity in the standards themselves. OSI International Standardized Profiles (ISPs) included a Protocol Interoperability Conformance Statement (PICS) which required implementors to list the features, facilities, services, options, and parameters their implementations supported. The GILS project team, however, needed to address additional requirements in the GILS Profile and came to view a profile as a way to subset and simultaneously integrate various protocols as well as requirements such as data content standards for the GILS Core data elements.

The ZIG had resisted the ISP and PICS orientation to static profiling for Z39.50. With Z39.50 increasingly incorporated into applications beyond the traditional library and bibliographic information environment for which it was initially developed, profiles for specific applications, however, are providing an acceptable mechanism to specify Z39.50 in these new uses (e.g., National Spatial Data Infrastructure, Museum Informatics). One of the contributions this project has made to research and development of the standard has been to identify and document the benefits of a profile, and more specifically, the issues and problems related to developing the GILS Profile (Moen and McClure, forthcoming).

GILS profile development (along with the development of the WAIS Profile) was the catalyst for the Open Systems Environment Implementors Workshop Special Interest Group on Library Applications (OIW/SIGLA) to develop an understanding of what comprised a Z39.50 application profile.[6] This group, some of whom worked on the GILS project team, concluded that a profile should be based on actual customer or user requirements. These requirements would be brought to implementors, and together the implementors and customers would work to develop an understanding of the requirements that would be supported by a profile.

6.1. What is a Profile

A profile is "a set of one or more base standards, and where applicable, the identification of chosen classes, subsets, options and parameters of those base standards, necessary for accomplishing a particular function (International Organization for Standardization/International Electrotechnical Commission, 1992, p. 2). Profiles are also referred to as "functional standards," "implementation agreements," or "specifications." Since open systems standards often include choices and options, profiles specify the values and parameters of a standard for an application or implementation to increase the likelihood of interoperability and interworking. A profile, according to these definitions, is a set of implementation agreements that guide implementors in applying one or more standards in a specific and limited context.

The research team broadened this definition for the GILS Profile to include not only the specifications for Z39.50 and other relevant standards in the application but also other aspects of a GILS conformant server that are beyond the scope of these standards. The GILS Profile provides the specifications for the overall GILS application relating to the GILS Core locator records (i.e., those resources maintained by the U.S. Federal government) and completely specifies the use of Z39.50 in this application. Using the GILS Profile in GILS implementations will facilitate interoperability of independently developed components of GILS. Further, in developing the GILS Profile, the project team was aware of the need to understand and address interoperability issues with the currently installed base of available implementation technology.

6.2. The GILS Profile

The GILS Profile includes the complete specifications of a subset of Z39.50 for use in the GILS application. The GILS Profile, in addition, specifies necessary characteristics of the GILS application that are outside the scope of Z39.50 including reference to other emerging, existing, or ad hoc standards (e.g., URIs, record interchange formats, and mappings between formats, respectively). Separate implementations will have an improved likelihood of interoperability and interworking when they conform to a common profile.

This first version of the GILS Profile focuses on the requirements for a GILS server operating in the Internet environment. Although the GILS Profile addresses GILS servers only, it is understood that clients have roles in the execution of information retrieval activities. GILS clients will be able to interconnect with any GILS server, and these clients will behave in a manner that allows interoperability with the GILS server. Clients that support Z39.50 but do not implement the GILS Profile should be able to access GILS records but with less than full GILS functionality.

The GILS Profile addresses many aspects of GILS (e.g., intersystem interactions and information interchange) but does not specify user interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality. Z39.50 does not address these either. Yet, considerations such as database structure or type of search engine used in specific GILS implementations may determine how well such systems perform (e.g., user satisfaction in retrieval results, response time, and efficient use of resources). Implementors of locators,

whether they are agencies or contractors, will be responsible for determining systems specifications in areas not addressed by the GILS Profile.

6.3. Lessons Learned from Developing the Profile

The GILS document presented an overview of GILS, including its objectives, service requirements, and core requirements. These requirements, however, were often described in general terms rather than in terms of specific functional requirements. A prerequisite for profiling to occur is a clear understanding — in as much detail as possible — of the actual functional requirements. Therefore, the project team proceeded to develop an interpretation and understanding of the high-level requirements presented in the GILS document. As a result, the team delineated the functional requirements that could be addressed by the GILS Profile. To accomplish this, the project team agreed upon a model of the system architecture that adequately described the GILS operation and information flows. This activity is documented in "Using Z39.50 in an Application for the Government Information Locator Service (GILS): A Background Paper."

Several lessons can be learned from the experience of developing the GILS Profile. First, since profiling was a new activity for the Z39.50 community, there were no clear precedents available to guide this work. The original project proposal called for a five month process during which the profile would be developed and implementations based on the profile would be developed. This schedule did not account for the time (given available resources) needed to develop consensus on the profile.

A second lesson from the GILS Profile project is that profile development time can be shortened if the user or customer requirements brought to implementors are clearly defined at a relatively detailed level. In the case of GILS Profile development, the project team spent considerable time understanding, interpreting, and designing a model of the system architecture for GILS. This was necessary since, as was noted above, the document on which the GILS Profile work was based provided a set of high-level requirements, and these needed to be further delineated so that specific choices could be made about using Z39.50 and other relevant or emerging standards. Therefore, developing a good profile (i.e., in the way the project team had redefined the GILS Profile as a system profile) requires a system architecture model that guides the specific choices about the manner in which the various standards will be utilized in an application.

While the general concept of GILS was relatively stable, some of the details shifted during the time of profile development. As an example, the number and extent of data elements to be used in locator records was in flux during the initial period of development work. Another lesson related to the one listed above is that the profiling process will proceed more rapidly if the specifications are stable and unchanging. However, this may be an unrealistic expectation, since following the model of software design where a customer and a software designer go through iterations of modelling, requirements delineation, etc., it is likely that any profiling process for a complex information system will need to be somewhat flexible and open to modification. As profile developers come to an understanding of what the requirements are, so also does the customer or user come to an understanding of the functionality the standard can support. In cases where certain requirements could not be met by the standard, requirements must be able to be modified.

A responsibility of the project team was to document the work of the project team and disseminate or make that information available to potential implementors and other stakeholders. The public dissemination via the Internet of drafts of the profile and the background document was intended to keep the process open and interested parties informed of the work of the project team. Providing information about the project's progress and requesting comments and responses made the process more open; participation was broadened since others external to the immediate project team could provide input into the decisions. Such public dissemination of information about the project also kept stakeholders who were not represented on the project team abreast of the directions in which the profile was moving. This meant that those who worked directly on the GILS Profile did not have an unfair advantage in developing implementations.

6.4. Impact on Product Development

The intention of the project leaders was to develop the profile, encourage test implementations based on the profile, and then have those test implementations provide feedback on the profile. The project leaders also intended to set up a mechanism by which these prototype implementations of the GILS Profile could undergo interoperability testing. The testing would

provide additional feedback to the project on the utility of the Profile, and if necessary, changes and/or expansions to the GILS Profile could be made. Time limitations precluded the development of test implementations, and therefore no interoperability testing occurred. Such testing, however, remains an essential component to profile development.

Testing would provide information such as validating whether or not the profile is implementable. Project leaders intended that the GILS Profile should enable implementations that would build on existing code base of Z39.50 implementors. Including actual implementors on the project team helped to ensure that the resulting profile could be characterized as "implementable." Discussions with Z39.50 implementors who reviewed drafts of the profile provided reassuring responses such as "this will require a low-level of effort to implement." In addition, testing could help to identify and correct unintended ambiguity or lack of precision in the profile.

Developing a profile that is relatively easy to implement should mean that off-the-shelf products can reach the market more quickly. This expands the choices to Federal agencies that will be implementing GILS servers. Client products supporting the GILS profile will be available and thus improve information retrieval for users.

6.5. Open Systems Environment Implementors' Workshop

The Open Systems Environment Implementors' Workshop (OIW) is sponsored by the Institute of Electrical and Electronics Engineers (IEEE) in cooperation with the National Institute of Standards and Technology (NIST). It is one of three international open systems workshops where implementors and other interested parties discuss and develop profiles supporting an open systems environment. In 1993, a group of Z39.50 implementors and others interested in the use of open systems standards for library applications established the Special Interest Group on Library Applications (OIW/SIGLA) to develop profiles related to Z39.50 and the international interlibrary loan standard.[7]

Project team members attended the quarterly OIW meetings (December 1993 and March 1994) and received continuing guidance on the development of the GILS Profile. Presenting drafts of the profile at the OIW meetings offered another opportunity for interested parties to comment and have input into the profile's direction and development.

At an interim OIW meeting held in conjunction with the April 1994 meeting of the ZIG, OIW meeting participants approved the GILS Profile. The OIW Plenary accepted the GILS Profile at its regularly scheduled June 1994 meeting. Upon acceptance at the plenary, the GILS Profile had the status of a "Working Implementation Agreement." According to the chair of the OIW/SIGLA, a vote on the GILS Profile as a "Stable Implementation Agreement" will be taken in September 1994. Attachment L is a copy of "Working Implementation Agreements for Open Systems Environment: Part 31 — Application Profile for the Government Information Locator Service (GILS) — Library Applications Special Interest Group," the official output from the June 1994 OIW as it relates to the GILS Profile.

6.6. The GILS Profile as a Federal Information Processing Standard

One of the ultimate goals of the GILS initiative is to have the GILS Profile accepted as a Federal Information Processing Standard (FIPS), which would then mandate its use across Federal agencies that are implementing locators. The GILS Profile, upon completion by the project team, was forwarded to NIST for preliminary processing as a FIPS. NIST placed a notice in the Federal Register on July 5, 1994 requesting comments on the proposed FIPS. This notice, "Proposed Federal Information Processing Standard (FIPS) for Application Profile for the Government Information Locator Service (GILS)" is included as Attachment M.

In summary, developing the GILS Profile provided an opportunity to explore new terrain of interest to the Z39.50 community. While learning specific lessons related to profiling needs and activities, the project has also contributed to the Z39.50 community's understanding of the usefulness of profiles. In addition, GILS Profile development contributed to the work of the ZIG in preparing a revised version of Z39.50, which is now being balloted through NISO. These contributions are outlined below.

7. Evaluation of Project

The plan for this project clearly identified a number of objectives.

- Expand research and development on the American National Standard for information searching and retrieval (Z39.50) for its application in facilitating public access to Federal information resources and speeding the development of interoperable systems
- Build consensus of major stakeholders on the manner in which Z39.50 can be applied in GILS implementations
- Develop an application profile for networked-based GILS implementations that references Z39.50 and other standards for use in the Internet environment
- Support and encourage test implementations of the profile by interested parties to provide evaluations of the profile and for interoperability testing.

This section reviews the extent to which the project achieved its objectives.

7.1. Consensus Among Stakeholders

The project team members responsible for developing the GILS Profile served as nominal representatives of various stakeholder communities, and the consensus building among team members was essential. In the meetings and the electronic discussions, team members arrived at a series of agreements about the GILS Profile that are reflected in the final product.

Contacting other potential stakeholders and interested parties for their response to the developing GILS Profile provided a means of establishing the validity of the project team's assumptions and agreements. While it is safe to assume that hundreds of people were alerted to the work of the project team and to the publicly available draft documents produced by the project, the actual amount of direct feedback from potential stakeholders and interested parties was not extensive. One reason for this low response rate can be attributed to the very technical nature of this undertaking. Without some understanding of the Z39.50 standard and the technical language used in the GILS Profile, many people were likely to find the documents too detailed, technically oriented, and thus not readily accessible to those who were not familiar with Z39.50. One of the respondents who contacted the project leader commented on the very technical nature of the documents.

More often than not, the comments received by project leaders and the project team, either in response to public presentations or to the documents, concerned issues raised by the GILS document rather than the use of Z39.50 within the GILS application. As an example, during a briefing on GILS conducted at the Coalition for Networked Information's Spring 1994 Meeting, one person asked about the meaning of the term "public" when qualifying information that would be described by GILS records. Another person questioned the adequacy of the data elements to capture information about information systems of interest to the archival and records management communities.

The project leader and project team made concerted efforts to remain in close contact with the ZIG, a primary stakeholder community. This group comprises technology providers, information services providers, university library systems and computing systems units, and others who are actually developing implementations using Z39.50. This group represents people who may develop GILS clients and servers, and thus it was essential that the members of this group support the work and results of the project team. At the Fall 1993, Winter 1994, and Spring 1994 meetings of the ZIG, the project leader provided updates on the work of the project, provided copies of the draft documents, and discussed with individual implementors their responses to the assumptions and agreements of the project team regarding the use of Z39.50. In addition, the electronic discussion list of the ZIG was used to announce available documents and request comments. While no comprehensive survey of the ZIG membership was undertaken, project leaders feel relatively confident that the responses received from ZIG members to the GILS Profile reflect a generally acceptable level of consensus.

A second important community contacted directly was the USMARC community. This was accomplished by working with representatives of the Network Development and MARC Standards Office, Library of Congress, on the proposed mapping of GILS data elements to USMARC fields. In addition, the project leader posted notices on the USMARC electronic discussion list regarding the proposed mapping and re-

questing feedback from USMARC experts on the choices made by the project team for the mapping. That feedback was used to clarify the mapping and provided additional information regarding the use of MARC records in the GILS application.

The project leaders received only one major objection to the choice of Z39.50 in the GILS application. The objection was based on a concern that Z39.50 would constrain the use of future technological innovations in information retrieval in GILS implementations. Since the GILS document served as the high-level requirements document for this project, and since it required the use of Z39.50 as the appropriate standard, this was not a decision that the project team could reverse. More importantly, it is not a decision that the project team would want to reverse. However, two project team members worked with the project leader to develop a response to this objection. Since it provides a good explanation why Z39.50 is an appropriate choice for GILS, it is included as Attachment N.

In summary, the project successfully built consensus among key stakeholders on the manner in which Z39.50 and other relevant standards would be used in GILS. The specifications included in the GILS Profile reflect this consensus. It is important to note, however, that in a technical undertaking such as profiling for the GILS application, consensus building among the potential stakeholder communities is carried out most effectively out when the stakeholders are knowledgeable about the technical details of the use of Z39.50 or other technical aspects of the application (e.g., the use of MARC records). Comments, suggestions, and a general willingness to respond to the work of the project team by members of the ZIG and members of the USMARC community were an indication of their participation in consensus building as well as an important validation of the resulting consensus.

7.2. Impact on Z39.50 Standards Development and USMARC Specifications

The GILS document directed that the participants in the development of GILS should work with the voluntary standards developers. The recently published OMB Circular A-119, "Federal Participation in the Development and Use of Voluntary Standards," (Office of Management and Budget, 1993b) directs agencies not only to use voluntary standards but also to work with standards groups in the development of standards. The GILS Profile project is an example of a productive collaboration where the government works within the voluntary standards development process.

Recent policy statements concerning the NII also discuss the government's involvement in the standards development process. For example, the Federal Information Infrastructure Task Force's Agenda for Action (p. 9) acknowledged the need for information technology standards to create an open and interoperable NII and suggested that the "Government can catalyze this industry-driven process by participating more actively in private-sector standards-writing bodies..." While there may be a number of leverage points the government can use to further the agenda of the NII (e.g., regulation, funding research and development projects), the GILS project serves as a good example in which the government, through a modest investment, supported and catalyzed standards activity to serve its own information technology standards requirements. Such targeted investment in standards development activities can have major pay-offs for the government and the NII initiative.

At the outset of the project, project leaders and project team members were uncertain whether or not Z39.50 would support all functions envisaged for the GILS application. The project team generally agreed, however, that basic GILS functionality could be supported by Z39.50 and changes or enhancements to Z39.50 required by GILS would be brought to the Z39.50 standards process (i.e., the ZIG). In the end, however, the project team found no GILS requirements needing to be addressed by changes to the standard. Yet, the GILS profile development has had an impact on Z39.50 in other ways.

As noted above, profiling Z39.50 for specific applications is a new activity within the Z39.50 community. This project has provided (along with the development of the WAIS Profile) a positive example of how profiling can be useful for applications using the standard. The GILS Profile will now serve as one model for future Z39.50 profiles in terms of both form and content.

Second, this is the first public application using a specific feature of Z39.50, the Generic Record Syntax (GRS-1), and the work done in developing the GILS Schema has had a direct impact on proposed revisions to the standard. The issues raised in developing the GILS Schema provided the basis for clarifying what a schema is, the relationship of tagSets, tagTypes, and tagpaths for nested data elements. In addition, a num-

ber of elements defined for the GILS tagSet have been incorporated in tagSet-G (elements in tagSet-G can be used by many different schemas).

Finally, GILS incorporates the use of Uniform Resource Identifiers (URIs) to improve interoperability and navigation in the Internet environment. URIs comprise a set of related standards for encoding resource location and identification information for electronic and other objects. Although not driving the creation of a Z39.50 Uniform Resource Locator (URL, which is a class of URIs), the use of URIs in the GILS Profile gave additional support for current efforts by a number of ZIG members to define a Z39.50 URL. Encoding resource locations and identification information in a Z39.50 URL will improve the seamless navigation among information objects and information servers envisioned for GILS.

The GILS project team also worked closely with the maintenance agency for USMARC (i.e., the Library of Congress) in the preliminary and final mapping of GILS data elements to USMARC. USMARC is an implementation of ANSI Z39.2, the American National Standard for bibliographic information interchange (American National Standards Institute, 1985). More importantly, a discussion paper developed by the maintenance agency proposed changes to the USMARC bibliographic format to accommodate special needs of GILS application using MARC records. "Proposal No. 94-9: Changes to the USMARC Bibliographic Format to Accommodate Online Systems and Services" is included in Attachment O. Specifically, the discussion paper proposed changes to USMARC that include:

- A new code in the MARC field 042, Authentication Source, to identify these MARC records as GILS Locator Records
- Incorporating fields from the Community Information Format into the Bibliographic Format to accommodate address and hours of service information of agencies, organizations, distributors, and points of contact for ordering and information about Federal information resources
- Changes in the 008 field for identifying when these Federal information resources are online information systems and services as opposed to documents and files.

By working within the standards development processes for Z39.50 and USMARC, the GILS project was able to address functional needs and specifications of the GILS application and maintain the standards-based orientation of the entire GILS initiative.

7.3. Opportunity for Generating Awareness and Stimulating Interest

This project, through the outreach efforts of its project team members and the project leaders, successfully brought GILS to the attention of a wide range of people and organizations. While the project's initial concern was to contact potential stakeholders and build consensus among them on the manner in which Z39.50 would be used in the GILS application, a valuable by-product of this activity was to generate increased interest in and awareness of the GILS initiative. Combined with the efforts by Eliot Christian in circulating the GILS document to generate awareness and understanding of the GILS initiative, the development of the GILS Profile helped people to understand that this Federal initiative would be brought into operation founded on a standards-based technology using Z39.50.

Section 5.2.3 outlined the various presentations, meetings, and other outreach efforts of the project in its attempt to build consensus. Publicly available documents in electronic formats were available (as well as more limited distribution in paper format). These efforts will help to build — in the longer term — a critical mass of potential users of GILS, a critical mass that can encourage Federal agencies to speed up their efforts in deploying GILS locators.

8. Issues Remaining for Future GILS Activities

This research project has laid the technical groundwork for the development of GILS servers and clients using Z39.50. The GILS Profile provides the necessary specifications for dealing with the technology. Yet, in the course of the research project, a number of issues arose regarding GILS. Although not necessarily specific to the scope and responsibilities of this research project, the Principal Investigators believe that they are important concerns that will need to be addressed as the GILS initiative progresses and, in fact, need to be addressed to ensure the viability and ultimate success of GILS.

8.1. Implementing GILS

GILS will provide information about Federal information resources only to the extent that Federal agencies — whether Cabinet-level, independent, or other agencies — actively support and participate in it. Agencies will need to take the initiative to make the necessary organizational, policy, and technology decisions to deploy their locators. Two primary issues are involved here. Agency decision makers and appropriate staff must be informed and educated as to the need for and utility of agency locators, and the agencies must be motivated to participate in GILS. In addition, a major component — in fact, the vital underpinning — of GILS will be the creation of GILS records that describe Federal information resources. The quality of data available in GILS will be directly related to the way agencies decide to implement the record creation component.

Compared with the situation at the time of the 1992 report (McClure, et al., 1992), Federal agency officials have become much more aware of the evolving networked environment and, specifically, the Internet. Numerous agencies now have direct Internet connections and are already using the network to provide access to agency information resources as well as disseminating information to the public. It is unclear, however, the extent to which agencies understand the GILS initiative and the specific details of the technology and standards that are the foundation of GILS. Until agency officials understand how GILS will operate and what GILS can deliver, they may be hesitant to move ahead expeditiously with deploying agency locators. An important step in realizing the promises of GILS is to develop a range of educational programs regarding how GILS may be implemented in each agency. Such a program is especially important for mid- and senior-level IRM staff, systems managers, librarians, and program officers.

Policy directives, regulations, and laws may motivate agencies to deploy GILS locators. The 1993 revisions to OMB Circular A-130, particularly Sections 8a(5), "Providing Information to the Public," and 8a(6), "Information Dissemination Management System" (Office of Management and Budget, 1993a, p. 36072) provide the policy basis for agencies to develop information locators to assist the public in locating "government information maintained by or for the agency." The Circular also directed agencies to "establish and maintain inventories of all agency information dissemination products," as well as developing "other aids to locating agency information dissemination products..." The additional revisions to the Circular published in July 1994 (Office of Management and Budget, 1994, p. 37912-7913) addressed the importance of deploying information technology to create an open systems environment and directed agencies to

Develop information systems in a manner that facilitates necessary interoperability, application portability, and scalability of computerized applications across networks of heterogeneous hardware, software, and communications platforms.

The Circular also reaffirms the policy of OMB Circular A-119 that directed agencies to use voluntary standards and Federal Information Processing Standards. The GILS initiative and the GILS Profile, which specifically details the use of voluntary standards to enable interoperable, networked-based information systems, support these policy objectives and provide agencies with a technical approach for implementing the policy.

Important policy and implementation questions remain, however, regarding the coordination and oversight of GILS implementation and maintenance. The forthcoming OMB Bulletin on agency information locators, however, will be an important adjunct to A-130, since the Bulletin can be the basis for more detailed guidance on agency responsibilities for GILS implementation, the use of the GILS Profile when implementing agency locators, and OMB's compliance and enforcement expectations. The Bulletin can also address issues related to government-wide coordination of the GILS. Once OMB releases the Bulletin, additional analysis on the adequacy of the guidelines and requirements will be possible.

Implementation of GILS needs to take account of other existing or emerging network-based information services under development by the Federal government. For example, the National Technical Information Service's (NTIS) Fedworld service and the electronic access service of the Government Printing Office (GPO) offer opportunities for cross-linkage with GILS implementations. There is a need to coordinate and articulate the relationships between GILS and other Federal information dissemination activities to avoid confusion and minimize redundancy.

Although policy directives may be place (e.g., OMB Circular and Bulletin), agencies may be slow to implement GILS until they recognize the tangible benefits

that GILS can offer. For example, the law that requires agencies to maintain inventories of their information resources has been less than effective. When agencies, however, understand that agency information locators can help them manage their information resources, they express more willingness to develop such mechanisms (McClure, Ryan, & Moen, 1992, McClure, et al. 1990).

Agencies will be motivated to participate in GILS through an understanding of the tangible benefits that will likely result. These benefits may take the form of increased control of agency information resources, cost-savings based on better management of those resources, and finally, increased support from agency constituencies based on improved "service to the citizens." Agency participation in GILS, then, will not automatically happen. A consciously developed program of education to help agencies understand what GILS is, as well as a combination of policy directives, budget review, and agency self-interest can be a basis for motivating agency participation.

8.2. Bibliographic Control in the Networked Environment

A number of writers have commented on the need to "organize" information in the networked environment (see for example Lynch & Preston, 1992; Moen, 1992). Without such organization, users will continue to have difficulty in finding information at the time they need it. Instead, the operating metaphor on the Internet is "surfing" rather than searching. In the paper-based environment of traditional libraries, librarians and other information professionals have developed principles and practices related to organizing information so that users can find that information. Broadly defined, bibliographic control is "The skill or art ... of organizing knowledge (information) for retrieval" (Svenonius, 1988, p. 88). Further, Wilson suggests that "Bibliographical control is a form of power, and if knowledge itself is a form of power, as the familiar slogan claims, bibliographical control is in a certain sense power over power, power to obtain the knowledge recorded in written form" (Wilson, 1968, p. 4).

One aspect of bibliographic control is the description of the information resource. the GILS record contains data elements that a record source can use for descriptive purposes. Another aspect of bibliographic control concerns the determination of what the information resource is about. The GILS record contains data elements for controlled and uncontrolled index terms that can indicate the topic or topics to which the information resource pertains.

A third aspect of bibliographic control, and one which GILS currently does not address is authority control. Authority control is the means to bring together all bibliographic records by a certain author (or in the case of GILS, agency or organization), with a certain title, or on a certain subject by establishing an authoritative "heading" or "name" that will be used consistently throughout an information system. For example, if a user was interested in finding all information resources by the Environmental Protection Agency listed in GILS servers, the user could search on "environmental protection agency." Yet, if some GILS records refer to the agency as "EPA" or "Environ. Protec. Agen.," these records may not be returned, and the user would not retrieve potentially relevant GILS records.

The GILS Core Elements include instructions to record creators to use particular forms of agency names (i.e., those found in the U.S. Government Manual). Unless the record creators understand the need to use a consistent form of an entity's (e.g., person, agency, department, office, etc.) name, they may not give appropriate attention to this prescription. The information must be consistent in an agency's GILS records, and there is a need for consistency across all GILS records. The library community has developed specific tools such as the National Name Authority File (maintained by the Library of Congress) to achieve consistency in the millions of bibliographic records used by libraries.

Federal librarians — professionals trained in the principles and practices of cataloging and bibliographic control — may have an essential role to play in the creation of GILS records. Consistency of the information in GILS records (i.e., authority control) as well as appropriate index terms describing the information resources (i.e., subject analysis) are necessary components of a high quality and useful GILS. Federal librarians could assist in creating adequate and useful GILS records.

The GILS record structure outlined in Christian (1994) and more fully described in the GILS Profile provides the name and semantics of the various data elements to be contained in the record. Lynch (1992) argues that information semantics in a distributed

computing environment must be addressed if the promise of networked information is to be realized. The GILS Profile's use of a common vocabulary for data elements and a common information structure of the records is an important step forward. Accuracy, completeness, currency, and consistency of data in the records, however, will be criteria by which the quality of the data can be evaluated. While the technology (i.e., Z39.50 clients and servers) may be able to process locator information, the end users will be badly served if the data is lacking in quality.

8.3. Accommodating the Installed Base of Technology

One of the design considerations in the development of the GILS Profile was that GILS servers should be accessible via a variety of existing clients. These clients include currently installed Z39.50 clients, Wide-Area Information Server (WAIS) clients, gopher clients, and Mosaic clients. Accommodating this installed base of technology immediately widens the potential user base of GILS. Similarly, other computer-to-computer communication protocols and applications (e.g., telnet, file transfer protocol, electronic mail) might also provide paths to GILS information.

The GILS Profile specifically addresses the use of Z39.50 and other standards for GILS servers. The Profile, however, goes beyond the specification of computer-to-computer protocols and also describes GILS record structure, semantics, and the content of locator databases. Based on an understanding of the semantics of the GILS records, a well-constructed GILS client, for example, may be able to interpret cross-references in GILS records to display browsable menus to users or it may use the unique record identification codes to compare various records and eliminate duplicates. A client that "understands" the predictable behavior of a GILS server can take advantage of the numerous features and functionalities specified in the GILS Profile.

Installed Z39.50 clients (i.e., those that have implemented Z39.50-1992) should be able to interact with GILS servers. These clients, however, will likely not be able to function as effectively as Z39.50 clients built to support GILS data. Bibliographic Z39.50 clients, (i.e., those that support information retrieval of bibliographic data), however, comprise an important subset of Z39.50 clients for GILS since GILS records are an instance of bibliographic data. These bibliographic clients currently can process MARC records, and since the GILS Profile requires GILS servers to pass GILS records in USMARC record syntax, the bibliographic clients should be able to access, retrieve, and process GILS data.

Another major group of installed clients are those implementing WAIS technology. Many of the installed WAIS clients are based on Z39.50-1988 with extensions specified in the Internet Engineering Task Force (IETF) RFC 1625, "WAIS over Z39.50-1988" (St. Pierre, et al., 1994; included in Attachment I). Implementors of WAIS are now moving toward compliance with Z39.50- 1992 based on the "WAIS Profile of Z39.50, Version 2" (also included in Attachment I).

Because of the widespread implementation of WAIS clients (specifically WAIS clients based on Z39.50-1988), FS Consulting prepared a paper that analyzed the requirements of GILS (as they were being developed by the project team in January 1994) in terms of the ability of WAIS implementations to interact effectively with GILS. Specifically the paper addresses how well WAIS could serve as an application tool for GILS. This paper, "Critical Review of WAIS as an Application Tool for GILS," is included as Attachment I.

Based on the analysis, the paper concludes that while the GILS Profile could be implemented using WAIS based on Z39.50-1988 (Attachment I, p. 11):

> there would be a number of problems which would prevent the achievement of an ideal implementation. Notably the direct lack of support for Local Control Numbers, Control Identifiers and Availability Elements, more formalized support for hierarchical menu browsing (and "well-known" searches) and the lack of support for structured records such as USMARC records and GRS records ... would make such an implementation technically feasible, but not ideal.

WAIS implementations based on the "WAIS Profile of Z39.50, Version 2," however, could "easily" implement the GILS Profile. However, a number of areas still need to be addressed including (Attachment I, p. 11):

> Availability Elements, more formalized support for hierarchical menu browsing (and "well-known" searches) and the "G" element set, support for SUTRS records and USMARC records. If these areas are addressed, then the WAIS-92 profile would be a suitable candidate for a GILS implementation.

Some providers of WAIS products (e.g., WAIS, Inc., and the Clearinghouse for Networked Information Discovery and Retrieval) have announced that they will support the GILS Profile in their products.

Determining the robustness of the interoperability between GILS products and the installed base, or for that matter with new products, will be an important consideration. There will be a need for some method of demonstrating interoperability of GILS servers and various clients. This topic is addresses below in Section 9.1.

8.4. Extensibility of GILS

The GILS, as it is described in the GILS document and specified in the GILS Profile, provides a baseline for initial implementations. In addition, the specification of GILS does not preclude the introduction of additional requirements that may enhance its utility and expand its usefulness. For example, current GILS data elements will accommodate the description of spatial data, and the addition of other data elements related to spatial data may increase the usefulness of GILS for locating spatial data resources.

Another area in which GILS can contribute to the management of Federal information resources is that of records management and archival activities. This area serves as one example of how current GILS specifications might be extended; it also provides an example to examine what procedures and processes would be appropriate to manage extensions to GILS.

As a basis for understanding an area in which GILS can be used, project leaders contracted with David Bearman to develop a paper discussing records management needs and how GILS might accommodate those requirements. His paper, "Requirements for Accommodating Information Systems Information and Records Management Needs within the Proposals for Government Information Locator Services (GILS) and its Z39.50 Application," is included as Attachment K.

The Federal Records Act requires agencies to make and keep records of the activity of the U. S. Federal Government for the period of the records' continuing value. The GSA and the NARA share with the creating agency the responsibility for maintaining these records. Agencies are mandated to submit all records created in the course of business for scheduling by the NARA. Bearman points out that this system of records management, reporting, and scheduling has a number of problems that are relevant to the GILS (Attachment K, p. 3):

- Agencies have no simple way to report to the GSA and NARA when new recordkeeping systems are created in order to get a schedule and disposition authority.

- Citizens have no comprehensive list of records created by the Federal government in the conduct of its business, despite the fact that all such records are required to be publicly accessible unless specifically exempted under limited clauses of the Freedom of Information and Privacy Acts.

- GSA and NARA have no way of using agency directories to records in order to fulfill their statutory responsibilities related to records management.

GILS could provide a means by which the operational requirements of records and archives management and the informational requirements of citizens are met. It is important to acknowledge again that while GILS serves the citizens as a locator service for Federal information resources, GILS can become an integral tool in the management of information resources. In this case, GILS can be an important addition to the tools available to GSA and NARA. According to Bearman, however, GILS can only fulfill its intended function if it is implemented with some modifications to the proposed scope, data content, and service definitions. He details these modifications in his paper. Neither the vision of GILS nor the system design as developed by the GILS Project Team precludes the technical modifications suggested by Bearman.

Important new uses of GILS will likely emerge as people become more aware of GILS and as implementations of GILS appear. To provide for systematically extending the utility of GILS will require a:

- statement of need and the context for use (e.g., outlining how GILS could be used for records management and NARA reporting requirements as a way of eliminating redundant systems within the Federal government to which agencies must report records)

- specification of what changes are needed to accommodate the articulated need (e.g., proposing a new GILS data element that would contain data for "Scheduled Disposition").

In addition, there is a need for a process by which the GILS Profile can be extended when changes and extensions to it are necessary. Although NIST will likely maintain the GILS Profile when it is adopted as a Federal Information Processing Standard (FIPS), there may be a need for a more informal mechanism (such as a subcommittee of an interagency group) that could receive proposals for changes/extensions and act upon these requests for changes. Such a subcommittee would not necessarily do the work of revising the Profile, but could contract to have the revisions developed and then pass them on to NIST for inclusion in revised versions of the GILS Profile. The OIW/SIGLA will also be a mechanism for making changes to the GILS Profile.

The project team envisioned GILS as a flexible service employing technology and standards that will accommodate new applications. Any extensions to the GILS Profile must account for an installed base of GILS implementations, and therefore controlled changes to the Profile are necessary. Arbitrary and unilateral changes in implementations of the Profile will be a disservice to the larger GILS effort since those changes may jeopardize interoperability and data sharing which are important goals of the technical specifications.

8.5. Federal Information Policy Issues

Technologically, GILS is an implementable solution for agencies to improve IRM and for people trying to locate government information resources. Yet the technology will provide only one of the necessary conditions for GILS to be implemented. Federal policy must be developed that addresses the broader questions of GILS. OMB is currently drafting a Bulletin that addresses GILS and the use of the GILS Profile by agencies deploying locators (the draft Bulletin should be available in late Summer 1994). Legislation and previous regulations already speak to the need for locators to help people find and access government information.

While policy instruments alone cannot and will not guarantee the success of GILS, the articulation of clear policy can provide another necessary condition for deploying GILS. The policy of OMB Circular A-130 regarding agency information inventories and locators as well as the need to create open and interoperable systems through the use of standards, as noted above, provide an important policy basis for GILS. OMB, as stated in A-130, provides "overall leadership and coordination of Federal information resources management within the executive branch" (Office of Management and Budget, 1994, p. 37914) and therefore has a critical role in the execution of the GILS initiative. The forthcoming OMB Bulletin on agency locators will hopefully specify OMB's role in GILS oversight and coordination. Circular A-130 enumerates agency responsibilities in information resources management, and specific sections, such as 8a(5) and 8a(6), detail agency responsibilities regarding locator information for their information dissemination products, but more specific guidance is necessary regarding OMB expectations of agencies and their implementation of GILS. The locator Bulletin needs to address the general question of "who is in charge" of GILS by detailing the specific responsibilities of OMB (e.g., enforcement) and individual agencies (e.g., compliance).

Federal policy must address the issue of motivating agencies to participate in GILS. While a policy statement by itself will not make GILS a reality, a policy that clearly articulates agencies' responsibilities and requirements regarding information locators is essential. Agencies must understand the benefits they will receive when they establish locators and must be willing to use the GILS Profile when they implement their locators. One might speak of enforcing compliance with the policy, but it may be more effective to identify incentives for agencies to participate in GILS.

8.5.1. OMB's Roles and Responsibility

OMB has played a key role in the past year in developing the vision statement for GILS (i.e., the GILS document). This was done as part of OMB's involvement in the Information Infrastructure Task Force. In addition, OMB's revisions of Circular A-130 (Office of Management and Budget, 1993a, 1994) addressed Federal information access and dissemination concerns, and directed agencies to help the public locate government information they maintain through the development of information inventories and other aids to locating agency information. The development of the GILS vision and the policy that supports such an

initiative are important contributions by OMB to improved Federal information resources management and the access to Federal information resources.

The formulation of a specific GILS policy through the forthcoming OMB Bulletin will articulate agency responsibilities regarding their information locators. This combination of policy requirements (i.e., A-130, the Bulletin) and the development of the means to carry out the policy (i.e., the technical specifications in the GILS Profile) should be an effective basis for guiding agencies. The challenge to OMB, however, is to coordinate agency involvement in GILS and to ensure that the GILS initiative can truly be a government-wide system for locating Federal information resources. In addition, OMB has a range of techniques (described in the 1994 release of A-130 Section 10, "Oversight," p. 37914) to evaluate each agency's information resources management and monitor agency compliance with the Circular. Therefore, OMB is in a position to address compliance of and participation by agencies in GILS.

8.5.2. GILS and IRM Roles and Responsibility

Since GILS can be an important component of managing information resources, IRM units will be a focus for GILS implementations. IRM staff, however, may need additional education and training about the GILS initiative and specific implementation issues. The GILS education program (see Section 10) will need to address how IRM units can use GILS for internal IRM and also for providing public access to Federal information resources. Providing a broader context for GILS use will assist IRM units and their agencies in understanding the benefits of participating in GILS. Agency budgets for GILS implementations must include funds for training and education.

IRM staff may need additional specific tutorials related to GILS. First, they may need training in Z39.50 and other standards (e.g., URIs, USMARC) used in GILS. They will also need guidance in identifying the information resources to be placed in GILS. Finally, they will need to understand the critical issues related to GILS record creation and data quality. Government-wide guidelines for creating GILS records can be the foundation for consistent and high quality GILS records.

A partnership between agency IRM and library units could be the basis for successful GILS implementations. IRM staff would bring technical expertise about information systems, networking, database management, etc., and the library staff would bring the technical expertise regarding bibliographic description, record creation, authority control, etc. Such a partnership would draw on the strengths of both IRM and library staff.

8.5.3. GILS and Federal Library Community Roles and Responsibility

GILS offers the Federal library community an opportunity to become participants in government information management, access, and dissemination activities. Additional efforts at policy and implementation levels, however, may be necessary to bring this important stakeholder group into the GILS initiative.[8] Many Federal libraries have a mission to serve the information needs of their agencies. Since GILS agency locators will assist agencies in their internal information management activities, libraries can benefit from successful GILS implementation. Librarians will be able to identify and locate not only their own agencies' resources but resources held by other agencies throughout the government. Federal librarians are important stakeholders in GILS.

Federal librarians can also bring specific technical expertise related to bibliographic control and cataloging to their agency's GILS implementations. Consistency of the information in GILS records (i.e., authority control) as well as appropriate index terms describing the information resources (i.e., subject analysis) are necessary components of a high quality and useful GILS. Librarians' experience in these bibliographic control activities will make them valuable participants in GILS implementations. Librarians can contribute to and benefit from a partnership with their agency's IRM units.

Federal librarians' participation and contribution will occur only if they take the initiative to become involved. This can be at the level of individual librarians making contacts with those involved in their agency's GILS implementations (or better yet, helping to initiate and shape the agency's GILS implementation). The Federal Library and Information Center Committee (FLICC) can also be an important stake-

holder in GILS by organizing and educating the Federal library community in GILS activities. FLICC, as with the individual librarians, will need to respond to GILS, but more importantly, can help shape GILS implementations; to do so will require FLICC to make the GILS initiative a priority for action.

One important area is in need of immediate attention and offers FLICC and the Federal library community an immediate opening for action. Government-wide guidelines for GILS record creation, especially related to important issues of bibliographic control, are needed. Drawing upon the expertise of Federal librarians, FLICC could assist in the development of these guidelines. This would also be an opportunity for forging partnerships between the Federal library community, the IRM community, and other parties interested in GILS implementation.

8.5.4. GILS and Archives and Records Management Community Roles and Responsibilities

The 1992 study (McClure, Ryan & Moen, 1992) examined how a government-wide information locator could serve the records reporting requirements of NARA. GILS offers the potential to NARA and the agencies to reduce reporting redundancy. To ensure the potential is realized, GILS records will need to accommodate all essential NARA-required information.

The Bearman paper (Attachment K) addresses a number of issues related to GILS and archival reporting requirements. An important consideration is what information resources GILS will include. GILS can be an important tool for managing information resources — documents, databases, records systems, etc. GILS is also a tool identifying, describing, and locating these resources. Thus, the more comprehensively agency holdings are included in GILS, the more value GILS will have — for internal information management purposes and for assisting people who are trying to locate government information.

8.6. Policy and Technical Standards

An essential element of the GILS initiative is the focus on a standards-based application of information technology in the Federal government. Choosing appropriate standards, integrating them in an application, and then encouraging compliance with the standards-based solution becomes critical when the goal is to achieve interoperability across a wide variety of implementations using heterogenous hardware and software.

Policy statements such as OMB Circulars A-130, "Management of Federal Information Resources" (Office of Management and Budget, 19994), A-16, "Coordination of Surveying, Mapping, and Related Spatial Data Activities" (Office of Management and Budget, 1990), and others call for the use of standards to promote an open systems environment, enable the transfer of data and resource sharing, and more generally indicate an increasing recognition of the role of standards in the use and flow of information. GILS provides a good example where standards serve as enabling tools for broader policy goals. In this case, the policy goals include providing citizens with access to Federal information resources and encouraging better information management by agencies.

Moen (forthcoming) argues that there is a link between information technology standards and broader information policy goals. He suggests that the time may be appropriate to call for an information technology standards policy for the Federal government. Such a policy may help move the Federal government towards the interconnected and interoperable horizon of the evolving NII. Attachment P, "Building a Policy for Information Technology Standards," outlines one approach to developing a policy. To summarize, there are three major components that must be addressed in building a standards policy: (1) a policy goal; (2) a framework for deploying standards; and (3) a management strategy.

- Determining a Policy Goal that Guides and Directs Standards Activities The 1994 revisions to A-130 direct agencies to deploy information technology and build information systems in the context of or on a migration path to an open systems environment (Office of Management and Budget, 1994). The emerging NII is assumed to be built as an open environment, connecting a wide array of information appliances across a variety of network and telecommunications paths. The recently released Report of the Federal Internetworking Requirements Panel (Federal Internetworking Requirements Panel, 1994; referred to below as the FIRP report) discussed the need for a goal, if not a framework, to guide agencies in their choice of standards-based technologies. These three examples point to the need for an overarching policy goal that articulates how

BEST COPY AVAILABLE

information technology standards will be used by the Federal government. A goal of "interoperability" or "open systems" addresses the technology. A broader goal that links interoperability, open systems, information management, and information policy objectives would be desirable.

- Adopting a Framework for Selecting Appropriate Standards An overarching policy goal of creating open systems may provide a baseline for standards choice, but if standards are to assist in achieving broader information policy goals, more specific guidance on selecting standards is required. One potential framework might be the information life cycle (Spring and Bearman, 1988). The information life cycle is a planning, evaluation, and oversight concept used in managing information resources (Hernon, 1994). This concept acknowledges that the stages of the life cycle are interrelated. A framework such as this could assist agencies to view their choice of particular standards to process or produce information as having potential impacts beyond the information's immediate use.

- Developing a Management Strategy that Assists Agencies and the Federal Government in its Standards Activities and Use An overarching policy goal and a framework for identifying appropriate standards must be accompanied by a management and organizational response to standardization. Recent writers have discussed the need for organizations to establish mechanisms to carry out standards oversight and coordination (Ritterbusch, 1990; Betancourt, 1993).

A government-wide management strategy will need to address: (1) mechanisms for sharing information about agency standards activities and for coordination among those who are actually deploying the technology (e.g., interagency groups); (2) guidance and oversight of agency standards programs (e.g., OMB budget review process); (3) coordination of standards implementation across Federal agencies (e.g., a single agency such as NIST or OMB or an interagency group such as the Interagency Committee on Standards Policy); (4) roles and responsibilities for specific agencies in standards education and training; and (5) a coherent and effective way to measure, evaluate, and achieve agency compliance with and use of standards.

Finally, a realistic management strategy for standardization should acknowledge the time-frame appropriate to standards. While there may be short-term benefits, it is in the longer term that benefits accrue; as the director of NISO points out, "standards are a long-term investment — both for development and implementation....We must not lose sight of the long-term benefits of standards, and we must make the long-term commitment to realize those benefits" (Harris, 1994).

The FIRP report acknowledges that an earlier policy for developing open systems in the Federal government (i.e., the Government Open Systems Interconnection Profile or GOSIP) has not been successful. Today's environment for networking, however, is much different from those days in the 1980s when GOSIP was initially formulated; robust networks exist (e.g., the Internet), networked-based protocol products are available (e.g., Z39.50 products), and the networks support real applications. The FIRP report recognizes that standards are critical components in the networked environment and are a means for achieving goals for Federal internetworking such as: fulfilling Federal mission needs, enabling interoperability, providing for software and hardware portability, and lowering costs.

A policy for information technology standards can guide future initiatives such as GILS. More importantly, such a policy can provide a context for agency participation in initiatives like GILS. Agencies will have a basis for understanding the need for standards-based solutions.

8.7. Evaluating the GILS Profiling Process and Products

The project team and project leaders, based on the feedback from stakeholders and the consensus building that occurred in the process of developing the GILS Profile, have a high level of confidence in the specifications that are included in the Profile. Until implementations based on the GILS Profile are built and people gain experience by working with it, however, little evaluation of the Profile is possible.

The broader context of this issue is evaluating the process used in developing the GILS Profile, the utility of the Profile, and the policy and implementation

efforts related to GILS (e.g., acceptance and implementation of GILS by individual agencies). For example, while the project team thinks it has broken new ground in profiling the GILS application, the process needs to be evaluated to see if it is ground that others can productively retrace.

An evaluation process is needed so that in 18-24 months the results of this project and the overall GILS initiative can be reconsidered and reevaluated. Such an evaluation can document the utility of GILS, the GILS Profile, and the process by which the Profile was developed. Any evaluation, however, will need to be preceded by addressing the other key issues identified above and giving attention to the "next steps" proposed in this report.

9. Next Steps to Realize the Promise of GILS

Assuming that the issues outlined in Section 8 are addressed and steps are taken to resolve the issues detailed there, there are, in addition, two important steps that should occur in order to move GILS from vision to reality. The GILS Profile is a vehicle by which the vision articulated by Christian (1994) has been made more technologically concrete, yet the following two inter-related steps can be catalysts for achieving widespread deployment of GILS.

9.1. A GILS Interoperability Testbed

A fundamental tenet of the GILS vision is that a wide variety of independently developed software for clients and servers can be used in GILS as long as that software is based on standards described in the GILS Profile. This means that agencies and users alike will need some reasonable assurance that these independently developed GILS components will work together in an interoperable manner. Interoperability of clients and servers, (i.e., the implementations work together effectively) was an assumption in the development of the GILS Profile; seamless navigation among disparate servers using Z39.50 will occur only if there is functional interoperability.

The GILS project team assumed interoperability as a design consideration, but in actual fact, separate implementations based on the GILS Profile cannot be assumed to be interoperable. Profiles, as noted previously, specify the values and parameters of a standard for an application or implementation to increase the likelihood of interoperability and interworking.

As part of this research project, the paper prepared by Preston and Lynch addressed the issues involved in ensuring interoperability of GILS products and implementations (Attachment J).

Preston and Lynch describe two approaches to testing implementations to ensure that they work together effectively:

- Conformance Testing — a single implementation is compared to the standard to be sure that the implementation does what the standard specifies
- Interoperability Testing — two or more implementations are tested directly against each other, with the standard used primarily as a reference to adjudicate problems and incompatibilities, and secondarily as a guide to the functions to be tested and the general behavior to be expected.

Based on the analysis offered by Preston and Lynch, the Principal Investigators concur with their recommendation that interoperability testing is the more appropriate path to take with GILS. This also follows the FIRP report's endorsement of interoperability testing as a pragmatic way to deal with this question of demonstrating interoperability.

Although a precise definition of interoperability may be hard to articulate, Preston and Lynch (Attachment J, p. 5) state that functionally, the meaning of interoperability is clear:

> components of a system such as GILS communicate with one another effectively, correctly, and provide the expected services to the user of a GILS client. In a very real sense, users don't care why components of a system like GILS fail to interoperate, or what component is at fault; while there can be many causes for failure, a successfully functioning operational system is clearly demonstrable to users. Further, users will view GILS as a totality; while there are a large number of standards and agreements involved in making the GILS work (each with its conformance and interoperability issues), users are only concerned that the entire constellation of standards, agreements and system components interoperate together effectively.

They describe how interoperability testbeds have been used successfully to carry out interoperability testing. In an interoperability testbed (Attachment J, pp. 5-6):

> a focused effort is made over a fairly short period of time to develop a number of implementations based on a set of standards that define a distributed system, and to experiment with using these implementations to interoperate with each other. It is important to recognize that while one of the primary purposes of a testbed is to explore interoperability issues, a testbed typically takes on a broader role as a large scale experimental prototype for validating a system design.

Continuing from their recommendation that interoperability testing "should be the keystone of any program to further the development of an interoperable base of GILS clients and servers" (Attachment J, p. 14), they state that "interoperability testbeds, in part as a way of developing a de facto core of well-known reference implementations and in part as a way of simply moving early implementations to a more mature state and ensuring their mutual interoperability, deserves careful consideration" (Attachment J, p. 15).

Interoperability testbeds, in addition to providing an arena for testing implementations, can serve other purposes. A GILS interoperability testbed can help to identify problems that arise from the interaction between different standards as described in the GILS Profile. A testbed also can provide feedback about the profile and may lead to improvements and changes to the profile. As important as these technical concerns, a testbed can serve as public demonstrations of the viability of a suite of standards. User communities, as well as potential implementors, can see how a distributed information system such as GILS actually works. This can be one aspect of stimulating the market for GILS.

9.2. Stimulating the Market

Throughout the development of the vision of GILS and in the development of the GILS Profile, there have been efforts to inform and create awareness of what GILS is and what GILS will enable. Now that the technical groundwork for GILS is laid and policy is being established, efforts should continue in the direction of stimulating a market for GILS. "Stimulating the market" includes developing demand by users for GILS services and developing a supply of products and implementations conforming to the GILS Profile that can be readily available to agencies to deploy agency-based GILS locators.

The appropriate role of the Federal government in stimulating the market, however, will likely take the form of motivating (and requiring by policy directive) agencies to participate in GILS. This in turn may create the demand for GILS products and implementations that will spur private-sector commercial enterprises to deliver off-the-shelf products. Since there is a tight connection between the demand for GILS products and services and their availability from private-sector suppliers, a mechanism that supports ongoing interaction between the Federal agencies and GILS product and services developers could be constructive. One mechanism could be the interoperability testbed; another mechanism could be OIW/SIGLA; additional mechanisms are possible.

There are several activities that can encourage a market of users of GILS. As noted above, an interoperability testbed can be an important public awareness tool. Preston and Lynch (Attachment J) point out the need to involve the user communities in a testbed project. These communities may include government documents librarians, Federal depository librarians, and others. Another important user community include the "intermediaries" described in the GILS document. These "users" will serve other groups of users by adding value to basic GILS records through repackaging, filtering, organizing, and other activities. They in turn will be developing their own markets for their GILS products.

Education and publicity about the GILS must continue. Agencies can play an important role in these activities. As they develop their GILS locators, they will want to alert their constituencies to the availability of their GILS service. The more proactive an agency is in promoting its GILS locators, the more likely it will create a demand not only for its GILS information but also for agency information resources to which GILS records points.

One step that the project leaders took to continue discussions around the GILS initiative and to promote awareness and to stimulate the market was to establish a public electronic forum on GILS. The Coalition for Networked Information has agreed to house the GILS Forum listserv, and one of the Syracuse project

leaders will serve as moderator. Information about subscribing to the GILS Forum is included in Attachment Q.

Another activity that may be particularly useful will be to "showcase" particular agency efforts in developing GILS locators. Recognition and acknowledgement of agencies that offer well-implemented and comprehensive GILS locators can alert users to the service. Rewards and incentives available to agencies for participating may help broaden the scope of Federal information resources available in GILS and may also help to motivate some of the more hesitant agencies to get involved. The idea is to raise the expectations of users (and expand the numbers of users) of what GILS will be able to offer them; those users will then expect and demand that Federal agencies participate in GILS by implementing locators that support the GILS Profile. At the same time, agencies should receive recognition for taking leadership roles in GILS implementations.

The other part of "stimulating the market" is to ensure that off-the-shelf GILS server and client products are available to agency implementors, as well as being available to the direct and intermediary users of GILS. Again, the testbed can be an important mechanism to get robust products developed. A GILS testbed can facilitate the overall development of a marketplace of GILS products. The knowledge gained by the testbed participants can be used by subsequent implementors to speed up development of products. Thus the benefits of a testbed goes beyond the important concerns of system validation and other technical, implementation problem-solving.

Organizations and companies that would like to do business with Federal agencies involved in GILS must have a sense that there is really a business opportunity in GILS. Whether these organizations and companies supply the agencies with server or client software or whether they develop an entire GILS package (including record creation, database loading, network connection, interface design, etc.) will depend on three things:

- Agencies' understanding of the value of GILS
- Agencies active participation in GILS
- The ability of the companies and organizations to identify the agencies that are ready to provide the market for these products and services, and to offer the agencies appropriate products and services.

An interoperability testbed can offer benefits to the agencies who are considering procuring GILS products, since the vendors can be asked to show how their products interoperate with those implementations that have been proven out through testbed activities.

Stimulating the market can lead to important "network externalities" for GILS participants, implementors and users alike. A network externality occurs when "one consumer's value for a good increases when another consumer has a compatible good, as in the case of telephones or personal computer software (Farrell & Saloner, 1985, p. 70). Interoperable components the of the GILS provide the "compatible goods." As additional participants are involved (particularly agency implementors), the reach and range of each new user is expanded.

Direct federal funding of GILS-related activities is another mechanism that could be considered. In addition there are a range of funding sources that might be tapped. For example, the Library Services and Construction Act (LSCA) has in the past provided grants to state and local libraries to enhance their services. Similar grants could be solicited to encourage library automation vendors to provide linkages between existing online public access catalogs (OPACs) and GILS. NII-related project funding could also be tapped to support GILS education initiatives as well as GILS technology demonstration projects. Such funding might be a suitable source for sponsoring the GILS interoperability testbed. A final source of funding that could be used for GILS activities is the depository library program. Funding selected and high impact pilot, demonstration, and other implementation projects would be extremely helpful to show a wide range of users (both inside and outside the Federal government) what GILS can offer.

A market for GILS will likely not to result simply from policy statements by OMB, GSA, or Congress. Those statements will be necessary, but not sufficient (as was demonstrated in the case of GOSIP). An interoperability testbed, the development of a market of users, and providing incentives and rewards to agencies, however, may all combine effectively to spark the market for GILS.

10. Final Recommendations

As a way of summarizing and concluding this report, Figure 3 provides a set of recommendations to help move the GILS initiative forward. The technical underpinnings and the vision of the GILS are in place, but those alone will not bring GILS to life. The following recommendations are painted broadly since additional study and research may be necessary to carry out these recommendations. Where possible, the recommendations include specific research activities that can assist in carrying out the recommendation.

- Strategically Place GILS as Part of Major Government-wide Initiatives: The National Information Infrastructure and the Re-inventing Government Initiatives: More important than any other recommendation is the need to promote GILS as a long-term solution to a variety of information-related problems facing the Federal government. GILS will be a component of the information infrastructure of the Federal government which is in turn a component of the National Information Infrastructure. GILS can be viewed as an enterprise-wide application (although deployed in a decentralized manner) addressing IRM, access, and dissemination responsibilities of individual agencies and the aggregate Federal government. GILS will use information technology to improve information management and contribute to more effective and efficient Federal government operations. OMB's forthcoming Bulletin on GILS is one opportunity for articulating a policy that recognizes this strategic role of GILS.

  As a component of the NII, GILS provides an opportunity to show how the Federal government can utilize the emerging NII to communicate with the public and enter a new era of access and dissemination of government information. GILS will be useful as a system for locating government information, but the vision of GILS (Christian, 1994) includes the use of GILS specifications for identifying and locating a wide range of network-based information held by the commercial sector as well as by the government. In addition, there is the intention for GILS to become a component of a Global Information Infrastructure (GII). The strategic use of GILS for identifying and locating information within the NII and GII should be acknowledged as a strength of the GILS initiative.

  Internally, GILS can assist Federal agencies in managing information resources by providing a unified and standardized mechanism to identify and describe information resources. GILS can also address mandatory reporting requirements of NARA. Agencies will have a robust vehicle for knowing what information resources are available to agency staff. Agency staff can use GILS to identify important information resources existing in other agencies, and thus it serves as a support mechanism for the affinity groups described in the National Performance Review (National Performance Review, 1993) and FIRP reports. GILS can be an important tool for managing information resources (by knowing what information resources exist) and can assist in the elimination of redundant reporting and tracking systems. These are long-range benefits that will accrue to agencies while at the same time providing overall benefits to the Federal government.

  GILS will also increase Federal agencies' capability to serve citizens by reducing the level of effort required for citizens to identify and locate important government information resources. GILS will affect how government takes care of its information re-

Figure 3. Next Steps for GILS: Recommendations

- Strategically Place GILS as Part of Major Government-wide Initiatives: The National Information Infrastructure and the Re-inventing Government Initiatives
- Develop a Program for Educating Federal Agencies about GILS
- Identify a Strategy for Stimulating a Market of GILS Products and Services
- Establish a GILS Interoperability Testbed
- Develop and Distribute GILS Record Creation Guidelines

sponsibilities — from cost-effective information resources management to improved productivity to service to the citizen.

Promoting the strategic role of GILS in a Federal government that is internetworked and interconnected with itself will likely run into familiar obstacles and barriers such as proprietary interests in agency information, a "we can do it ourselves" attitude, and worries that increased citizen access to Federal information will mean more work for the agency. In addition, the power of the decentralized nature of the GILS may increase the difficulty of taking such a vision across the entire government. The tools may be conducive to decentralized deployment, but policy regarding how GILS will be deployed (e.g., requirements, data quality, utility) will be top-down.

A strategic stance regarding GILS deployment, maintaining visibility of GILS through activities and policy, and the use of highly-placed supporters of the NII and the re-inventing government initiatives to exhort agencies about the merits of GILS will be necessary to set the broader context for the Federal government's use of GILS.

- Develop a Program for Educating Federal Agencies about GILS: At this point it is unclear the extent to which agencies understand what GILS is, how it can work, and the benefits that can accrue to agencies participating in GILS. Policy statements and directives are one method of creating awareness within the agencies. This recommendation, however, is intended to reach out to the agencies and move from awareness to understanding.

  It is necessary, however, to point out that there is a general problem with educational programs that support Federal IRM (McClure, forthcoming). Since GILS is likely to be executed as a component of Federal IRM, some of the same key issues for an IRM educational program identified by McClure will likely be issues for GILS-specific educational activities. These issues include: what are the educational and training needs; who is responsible for coordinating educational and training programs; determining adequate funding mechanisms and funding sources; and how can sharing of innovations and knowledge be rapidly diffused across government.

  For GILS, an educational program is needed that can address broad policy and implementation implications such as identifying how GILS fits into the larger context of the government component of a NII and how GILS can improve agency IRM. At another level, the educational program can provide agencies with hands-on tutorials and demonstrations of how GILS implementations work. An educational program should address the GILS concept, real-world applications and use of GILS, technical demonstrations, and tutorials on how to implement GILS, including the identification of information resources to describe, creation of GILS records, and maintenance of GILS services. Finally, agencies need to develop a positive attitude and the necessary skills in "marketing" their information products and services. Thus, the educational program for GILS needs to be broad-based and cover a variety of training and education needs.

  Funding a GILS educational program must be addressed. Agencies must anticipate training and education expenses when developing their plans for agency locators. Government-wide and interagency training are possible, and possible government-wide funding should be explored (e.g., through NII-related programs and projects).

  Interagency groups such as the Federal Information Resources Management Policy Council (FIRMPoC) or the Interagency Working Group on Public Access (the "Solomon's Group") could coordinate a GILS education program. This could be done by drawing on the expertise and resources of agencies that are leaders in GILS implementations. Establishing partnership with private sector vendors who are developing and marketing GILS products and services is also possible. These vendors could be important sources of expertise on Z39.50 and GILS, and vendor-led GILS seminars and training could be cost-effective mechanisms for components of the GILS educational program.

  The primary point here, however, is that the work on GILS to date (i.e., vision statement, technical specifications, policy instruments) are not the final step; this work has only laid the groundwork. Additional steps will be necessary to assist agencies in developing their GILS locators.

- Identify a Strategy for Stimulating a Market of GILS Products and Services: As one part of a GILS educational program, efforts are needed to stimulate a marketplace for GILS products. The GILS vision includes a partnership with non-government organi-

zations and businesses (e.g., value-added suppliers of government information, vendors of technology products). There is little reason for agencies to think that they need to build GILS implementations from scratch since commercial off-the-shelf products may well be available from vendors. Yet the agencies need to be convinced of the central role they play in seeing that GILS servers are deployed that identify and describe their information resources.

The development of the GILS Profile brought together Z39.50 implementors who are interested in supporting the GILS Profile in their implementations. What is needed, however, is a strategy to motivate agencies to: 1) participate in GILS by deploying GILS servers, and 2) to encourage the agencies to procure off-the-shelf GILS products whenever possible. One of the motivations for a standards-based solution offered by the GILS is to rely on private sector vendors of robust Z39.50 GILS implementations — clients and servers — to provide the technological components of the agency GILS. The private sector will support GILS through the development of products and services only to the extent that businesses see a commercial opportunity. Vendors may incorporate support for the GILS Profile into their Z39.50 client implementations, but unless agencies deploy GILS servers those clients will serve little purpose.

- Establish a GILS Interoperability Testbed: An interoperability testbed can serve as an adjunct in stimulating the development of GILS products (e.g., GILS clients and GILS servers). Such a testbed is necessary also for identifying any shortcomings in the GILS Profile so that it may be corrected sooner rather than later. Even two or three participants in an early informal testing environment would be helpful to validate the overall GILS system design as well as the specifications in the Profile.

  The testbed would also provide implementors with experience in developing robust GILS products, and the testbed may reduce the time needed to move from the development of products to their availability in the market. Sharing the results of the testbed is essential, and to provide a suitable environment for information sharing, the testbed should likely be sponsored by a neutral third-party. For example, the Coalition for Networked Information sponsored the 1992 Z39.50 interoperability testbed.

  A GILS testbed, due to the nature of the GILS specifications as Preston and Lynch (Attachment J) point out, will likely be a more complex undertaking than other testbeds since such a testbed will be testing the whole system — from server behavior to database semantics. Thus, preliminary to establishing a testbed, this recommendation suggests the need for a research project be conducted to examine issues that may be critical factors in the success and utility of the testbed. Preston and Lynch provide a beginning list of such issues.

- Develop and Distribute GILS Record Creation Guidelines: An important factor in the overall utility of the GILS will be the quality of the data in GILS records. Quality criteria will include accuracy, consistency, completeness, and currency. In order to encourage the creation of high quality information that will populate GILS servers, the development of written guidelines for creating GILS records is essential.

  The library community has a long history in identifying and describing information. Librarians have developed very specific principles and standards for creating bibliographic records. It is likely that the Federal library community can play an important role in the generation of GILS records. For example, the Library of Congress (LC) has coordinated cooperative name authority work for many years and maintains a national name authority file. The National Library of Medicine (NLM) and the National Agricultural Library (NAL), as well as LC, employ experts in cataloging and bibliographic control. This expertise at the Federal library level could be coordinated by FLICC for developing the GILS record creation guidelines. An important benefit of utilizing these experts is their familiarity with the problems of bibliographic control, the solutions to these problems, and the use of MARC in automated library systems.

  Any guidelines for GILS record creation should be developed incrementally so that it is clear that the guidelines are offered as a way of solving problems that arise in how to describe a Federal information resource. Such guidelines should not attempt to address all possible circumstances, but rather provide general guidance on how to solve these problems.

  These guidelines should be developed with the assistance and participation of creators of the records and the users of the records. This recommendation

could be carried out in reasonably short-order through a small research project. Developing these guidelines needs to be done quickly. Agencies are beginning to create GILS locator records, and they will need guidance sooner rather than later on the data and their format that go into GILS records.

11. Moving Forward

As described in this report, much progress has been made in translating the GILS concept into implementable products. This most recent effort undertaken by the Principal Investigators concentrated on standards development, consensus building among key stakeholder groups, and interoperability among GILS components that are likely to evolve from the Federal agencies. But our work in this area over the last six years suggests that there are a number of factors that must be kept in mind if additional progress is to be made on developing and implementing GILS in the Federal government.

Figure 4 offers an overview of some of the key factors that will effect future development of the GILS. Some of these factors have been previously discussed in this report; others have been identified in previous studies by the authors (McClure et al., 1990; McClure, Ryan, and Moen, 1992). Receiving inadequate attention, thus far, is the social/cultural organizational factors in Federal agencies that prescribe a set of attitudes and behaviors that will either support or mitigate against GILS development.

To date, we have found a wide variance in the cultural context among the agencies as to their attitudes and behaviors regarding:

- importance of providing public access to government information
- organizing the agency to facilitate that access to government information
- knowledge and implementation of laws and regulations related to managing information technology, standards, and public access
- obtaining ongoing feedback and assessment from the agency's primary customers regarding the agency's information services and products.

In short, strategies and policies from OMB-OIRA, for example, regarding GILS development will have to be tempered by different situational contexts among the

Figure 4. Key Factors Affecting GILS Implementation

agencies. Thus, flexibility in policy development will be essential. Some agencies are only just understanding the concept of a GILS while others are beginning to deploy them.

An important lesson resulting from this most recent project is that consensus building among key stakeholders such as Federal agencies, information services and products developers, researchers, and users is essential. In this particular instance, stakeholders looked for solutions to solve interoperability issues and to define Z39.50 applications for GILS. This context of problem solving and moving forward is an important result of the project. We hope that such an approach will continue in the future as GILS become implemented and used.

In 1988 when we began work in the area of a government-wide information inventory locator system, we received minimal interest from government agencies and information providers. The early support, first from OMB-OIRA and next from the Center for Electronic Information at NARA, was instrumental in moving from the concept of a GILS to real-time systems such as we have today. But while much progress has been made over recent years in implementing an interoperable GILS throughout the Federal government, much remains to be done. The recommendations offered earlier in this report suggest which next steps should be addressed.

Improved public access to Federal information and improved management of that information by individual agencies has been a policy goal clearly articulated by the Clinton Administration. The efforts to develop and implement GILS are important means to accomplish this policy goal. As this report suggests, significant progress has been made, now, on the technical side for implementing and ensuring interoperability for GILS. Future emphasis on GILS development will need to shift to individual agencies and to focus on encouraging cultural and organizational changes needed to support GILS implementation and use.

Notes

1. Uniform Resource Identifiers (URIs) is a generic term referring to a set of related standards for encoding resource location and identification information for electronic and other objects. The URI Working Group of the Internet Engineering Task Force (IETF) defines and specifies URIs. There are currently three objects within the URI set: the Uniform Resource Locator (URL); the Uniform Resource Name (URN); and Uniform Resource Characteristics (URC). The URI Working Group has approved URLs for experimental standardization, and it is expected to approve URNs in 1994. URCs are in the developmental stages.

2. Z39.50-1992 (also referred to as Version 2) was approved in 1992. Since that time, the Z39.50 Implementors Group (ZIG), which is a voluntary user group comprising implementors of Z39.50, has continued work to enhance the standard based on needs of information providers and users. Balloting on a new version of the standard, Z39.50-199x, began September 1, 1994. During the development of the new version, it was sometimes referred to as Version 3. According to Ray Denenberg, the editor of the standard and member of the Z39.50 Maintenance Agency at the Library of Congress, "it is important to point out, however, that although these version designations do have specific protocol significance, they should not be used to refer to versions of the standard. Z39.50-1992 specifies protocol version 2, and Z39.50-1994 specifies protocol versions 2 and 3." A draft of Z39.50-199x can be retrieved from the Library of Congress's gopher. Connect to MARVEL.LOC.GOV and select #7. Services to Libraries and Publishers, and then select #8. Z39.50. Or via anonymous FTP at <ftp.loc.gov> in the directory /pub/z3950.

3. The acronym GIILS refers to the Government Information/Inventory Locator System concept proposed in the 1990 and 1992 research projects (McClure, et al., 1990; McClure, Ryan, and Moen, 1992). While there are many similarities between the GIILS and the Government Information Locator Service or GILS, readers should note that these are not equivalent. References to GIILS reflect the understanding of the researchers in 1990 and 1992. Our understanding of such a locator system has evolved through the vision presented in the GILS document and the work on this research project.

4. To give some indication of the number of people reached through the postings to these electronic listservs, the following gives the approximate number of subscribers for each (as of August 1994): CNI-ANNOUNCE — 915; GOVDOC-L — 2,250; PACS-L — 8,150; USMARC — 670; Z3950IW — 475. In some cases the individual "subscriber" is mail re-

flector/exploder, electronic bulletin board, or Usenet reader, and thus the actual number of people who received or read these posting is likely much higher than the sum of the numbers listed.

5. Another profile was under development at the same time as the GILS Profile. Work on the WAIS Profile has been done by representatives of WAIS, Inc., Clearinghouse for Network Information Discovery and Retrieval, the Library of Congress, and members of the Open Systems Environment Implementors' Workshop Special Interest Group on Library Applications.

6. The OIW is the forum where implementors come together to develop implementation agreements. The Special Interest Group on Library Applications (OIW/SIGLA) was formed to develop profiles for Z39.50 and other open systems standards related to library applications (e.g., the international standard for interlibrary loan).

7. "Workshop Policies and Procedures" (available from NIST) provides information about the OIW and its procedures. A charter of the OIW/SIGLA is available from the chair of the SIG; contact Ralph LeVan, Senior Consulting Analyst, OCLC Online Computer Library Center, Inc., Mail Drop 432, 6565 Franz Road, Dublin, OH 43017; email: <rrl@oclc.org>; phone: 614-764-6115.

8. The project leaders contacted members of the Federal library community in the course of the study and requested input (in the form of a contributed paper to the study) on the roles and responsibilities of the Federal library community in GILS. Conversations with individual Federal librarians pointed out to the project leaders some of the concerns of the librarians. While these librarians expressed interest in participating in GILS and helping to shape and contribute to GILS, the project leaders received no formal response from the Federal library community regarding GILS.

References

American National Standards Institute. (1985). American national standard Z39.2-1985: Bibliographic information interchange. New York: American National Standards Institute.

Betancourt, Diego. (1993, November 19). Strategic standardization management: A strategic macroprocess approach to the new paradigm in the competitive business use of standardization. [Available from the author: Diego Betancourt, Manager, Office of Strategic Standardization, Polaroid Corporation, 656 Technology Square, Cambridge, MA 02139]

Christian, Eliot. (1994, May 2). Government information locator service (GILS): Report to the Information Infrastructure Task Force. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at <130.11.48.107> as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

Clinton, William H. & Gore, Albert, Jr. (1993, February 22). Technology for America's economic growth: A new direction to build economic strength. Washington, DC: Government Printing Office.

Farrell, Joseph and Saloner, Garth. (1985, Spring). Standardization, compatibility, and innovation. Rand Journal of Economics 16(1): 70-83.

Federal Internetworking Requirements Panel. (1994, May 31). Report of the Federal Internetworking Requirements Panel: Prepared for the National Institute of Standards and Technology. Gaithersburg, MD: National Institute of Standards and Technology. Available via anonymous FTP at <osi.ncsl.nist.gov> in directory /pub/firp/

Harris, Patricia Harris. (1994). National standards development in the United States: Private sector interests in the public sector environment. In Scholarly Information and Standardization, edited by William E. Moen (Bethesda, MD: NISO Press, 1994), p 32.

Hernon, Peter. (1994). Information life cycle: Its place in the management of U.S. government information resource. Government Information Quarterly, 11 (2): 143-170.

Information Infrastructure Task Force. (1993, September 15). The national information infrastructure: Agenda for action. Washington, DC: NTIA NII Office, Department of Commerce. Available in ASCII text format, filename: niiagenda.asc on the NTIA Bulletin Board (202-482-1199) and the Fedworld bulletin board (703-321-8020); available by anonymous FTP from

<ftp.ntia.doc.gov> as /pub/niiagenda.asc; available by gopher at <gopher.nist.gov> in the menu item DOC Documents.

International Organization for Standardization/International Electrotechnical Commission. (1992). ISO/IEC TR10000-1 Information technology — Framework and taxonomy of international standardized profiles — Part 1: Framework. Geneva: ISO/IEC Copyright Office.

Lynch, Clifford A. (1992). Online searching on the internet: The challenge of information semantics for networked information. In Proceedings of the 1992 National Online Meeting (pp. 203-210). Medford, NJ: Learned Information.

Lynch, Clifford A. (1994a, April). Using the Z39.50 information retrieval protocol in the internet environment. Information Standards Quarterly 6(2), 1-5

Lynch, Clifford A. (1994b, August 23). Using the Z39.50 Information Retrieval Protocol in the Internet Environment [Draft RFC]. Available via anonymous FTP at <ds.internic.net > as /internet-drafts/draft-ietf-iiir-z3950-00.txt

Lynch, Clifford A. & Preston, Cecilia M. (1992, Spring). Describing and classifying networked information resources. Electronic Networking: Research, Applications and Policy 2(1): 13-23.

McCallum, Sally H. (1994). Information technology standards: Implementation, maintenance, and coordination. In William E. Moen (Ed.), Scholarly Information and Standardization (pp. 13-20). Bethesda, MD: NISO Press.

McClure, Charles R. (forthcoming). Promoting education for information resources management in the federal government. Government Information Quarterly.

McClure, Charles R., Bishop, Ann, Doty, Philip, & Bergeron, Pierrette. (1990). Federal information inventory/locator systems: From burden to benefit: Final report. Syracuse, NY: Syracuse University, School of Information Studies. (ERIC Document Reproduction Service ED 326-247)

McClure, Charles R., Moen, William E. & Ryan, Joe. (Winter, 1992). Design for an internet-based government-wide information locator system. Electronic Networking: Research, Applications and Policy 2(4): 6-37.

McClure, Charles R., Ryan, Joe & Moen, William E. (1992). Identifying and describing federal information inventory/locator systems: Design for networked-based locators. Capitol Heights, MD: National Audio Visual Center. (ERIC Document Reproduction Service ED 349-031)

Moen, William E. (forthcoming). Information technology standards: A component of federal information policy. Government Information Quarterly [forthcoming].

Moen, William E. (1994). The Z39.50 protocol: Information retrieval in the information infrastructure. Bethesda, MD: NISO Press.

Moen, William E. (1992). Organizing networked resources for effective use: Classification and other issues in developing navigational tools. In Networking, Telecommunications, and the Networked Information Revolution, Proceedings of the ASIS 1992 Mid-Year Meeting, Albuquerque, NM, May 28-30, 1992. Silver Spring, MD: American Society for Information Science, 1992.

Moen, William E. and McClure, Charles R. (forthcoming). A profile for the government information locator service (GILS) using ANSI/NISO Z39.50: Implications for a user-based approach to standards development. StandardView.

National Information Standards Organization. (1992). ANSI/NISO Z39.50-1992, Information retrieval application service definition and protocol specification for open systems interconnection. Gaithersburg, MD: NISO Press.

National Institute of Standards and Technology, Computer Systems Laboratory. (1994, July). Framework for national information infrastructure services [Draft]. Gaithersburg, MD: U.S. Department of Commerce, National Institute of Standards and Technology, Computer Systems Laboratory.

National Performance Review. Office of the Vice President. (1993, September), Reenginering through information technology: Acccompanying report of the national performance review. Washington, DC: Government Printing Office.

Office of Management and Budget. (1994, July 25). Management of federal information resources [Circular No. A-130]. Federal Register 59: 37906-37928.

Office of Management and Budget. (1993a, July 2). Management of federal information resources [Circular No. A-130]. Federal Register 58: 36038 - 36086.

Office of Management and Budget. (1993b, October 26). Federal participation in the development and use of voluntary standards [Circular No. A-119]. Federal Register, 58: 57643-57648.

Office of Management and Budget (1990). Coordination of surveying, mapping, and related spatial data activities [Circular No. A-16].

Ritterbusch, Gerald H. (1990). Functions of a standards program. In Standards Management: A Handbook for Profits, edited by Robert B. Toth (New York: American National Standards Institute, 1990), pp 19-34.

Spring, Michael B. and Bearman, Toni Carbo. (1988). Information standards: Models for future development. Book Research Quarterly, 4: 38-47.

St. Pierre, Margaret, et al. (1994, June). RFC 1625, WAIS over Z39.50-1988. Available from the InterNIC gopher. Connect to <ds.internic.net> and select: 4. InterNIC Directory and Database Services/ 6. Internet Documentation (RFC's, FYI's, etc.)/ 7. RFC's (Request For Comments).

Svenonius, Elaine. (1988, January/March). Directions for research in indexing, classification, and cataloging. Library Resources & Technical Services 25 (1): 88-103.

Wilson, Patrick. (1968). Two kinds of power: An essay on bibliographical control. Berkeley and Los Angeles: University of California Press.

APPENDIX A
Project Team Members

The research project team consists of experts in Z39.50 and also representatives of Federal agencies. The following lists these members:

Z39.50 Experts

Kevin Gamiel	Clearinghouse for Network Information Discovery and Retrieval
Ralph LeVan	OCLC
Denis Lynch	ESL, Inc.
Margaret St. Pierre	WAIS, Inc.
Madeleine Stovel	Research Libraries Group, Inc.

Representatives from Federal Agencies

Eliot Christian	United States Geological Survey
Tim Gauslin	United States Geological Survey
Sue Ruddle	Defense Technical Information Center
Yesha Yelena	National Institute of Standards and Technology

Representative of the Z39.50 Maintenance Agency

Ray Denenberg	Z39.50 Maintenance Agency, Library of Congress

Attachment A

EXPANDING RESEARCH AND DEVELOPMENT ON THE NISO Z39.50 SEARCH AND RETRIEVAL STANDARD

PROJECT ABSTRACT

Charles R. McClure
William E. Moen
School of Information Studies
Syracuse University, Syracuse, NY 13244
Phone: 315-443-2911
FAX: 315-443-5806
Email: CMCCLURE@SUVM.ACS.SYR.EDU WEMOEN@RODAN.ACS.SYR.EDU

October 1, 1993

Agencies of the Federal government increasingly use information technology to create, process, store, and disseminate information in digital form. Recent policy decisions such as the revised Office of Management and Budget Circular No. A-130, "Management of Federal Information Resources," direct agencies to make public information available in electronic formats and to develop the mechanisms by which they can disseminate this information using the emerging national information infrastructure.

One major barrier to effective citizen access to public electronic information is the lack of directories and other finding tools to determine information resources Federal agencies hold and disseminate. A recent project by the authors ("Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators" by Charles R. McClure, Joe Ryan, and William E. Moen) concluded that an agency-based, network-accessible Government Information Locator could assist users to access and agencies to disseminate public information.

To advance the development of this Government Information Locator Service (GILS), the Interagency Working Group on Data Management for Global Change is funding a cooperative agreement between the United States Geological Survey and Syracuse University, School of Information Studies. McClure and Moen, co-principal investigators, are coordinating a research project focused on the use of open systems standards to improve the utility of information searching and retrieval on digital networks. More specifically, the project has as its objectives to:

- Expand research and development on the American National Standard for information searching and retrieval (ANSI/NISO Z39.50) for its application in facilitating public access to Federal information resources and speeding the development of interoperable systems
- Build consensus of major stakeholders on the manner in which ANSI/NISO Z39.50 can be applied in GILS implementations
- Develop an application profile for a networked-based GILS implementations which references ANSI/NISO Z39.50 and other standards for use in the Internet environment
- Support and encourage test implementations of the profile by interested parties to provide evaluations of the profile and for interoperability testing.

The project will bring together a research team of experts on ANSI/NISO Z39.50 and Federal information resources to carry out the objectives of the project. The project began September 7, 1993 and is scheduled for completion in February 1994.

A variety of implementors now use ANSI/NISO Z39.50 in applications for networked-based information searching and retrieval, and this project will build on the experience and expertise available from these early implementors. Federal agencies view this standard as providing the basis for locator applications.

Agency staff have outlined the functionality needed for the GILS, and the profile development that is the goal of this project will address these user requirements and at the same time provide specific guidance for implementations of the GILS. Constructing a standards-based GILS, based on a widely accepted application profile, will ensure interoperability and interworking of the agency implementations. As important, the standards-based GILS will ensure wider access to Federal information resources.

A summary report of the project will be available and will document the activities, products, and accomplishments of the project.

Attachment B

APPLICATION PROFILE FOR THE GOVERNMENT INFORMATION LOCATOR SERVICE (GILS)

Developed as Part of the Cooperative Research Study:

Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard

between the
School of Information Studies, Syracuse University
and
The United States Geological Survey
funded by
The Interagency Working Group on Data Management for Global Change

Charles R. McClure <cmcclure@suvm.acs.syr.edu>
William E. Moen<wemoen@mailbox.syr.edu>
Co-Principal Investigators

School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244-4100
Telephone: (315) 443-2911
Fax: (315) 443-5806

May 7, 1994

CONTENTS

1. Introduction

2. Background

3. Scope

4. Field of Application

5. References

6. Definitions

7. Z39.50 Specifications for GILS

7.1. Version
7.2. GILS Objects
7.3. Communication Services
7.4. Z39.50 Services
7.4.1. Search
7.4.1.1. Attribute Set
7.4.1.2. Well-known Search
7.4.2. Retrieval
7.4.2.1. Schema
7.4.2.2. Element Sets
7.4.2.3. Record Syntax
7.5. Preferred Display Format
7.6. Diagnostic Messages

8. GILS Core Elements

Annexes

Annex A — GILS Attribute Set
Annex B — GILS Core Element to USMARC Mapping
Annex C — Preferred Display Format for GILS Records
Annex D — GILS Schema
Annex E — GILS Core Elements

APPLICATION PROFILE FOR THE GOVERNMENT INFORMATION LOCATOR SERVICE(GILS)

1. Introduction

This document describes an application profile for the Government Information Locator Service (GILS). The GILS Profile includes not only the specifications for ANSI/NISO Z39.50, the American National Standard for Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (National Information Standards Organization, 1992) in the application but also other aspects of a GILS conformant server that are outside the scope of Z39.50. The GILS Profile provides the specifications for the overall GILS application relating to the GILS Core, which is a subset of all GILS Locator Records, and completely specifies the use of Z39.50 in this application.

2. Background

The GILS is a response to the need for users to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources. Christian (1994) is the authoritative document providing an overview of GILS, its objectives, service requirements, and core requirements. According to Christian (1994), the GILS is an overall service and includes information and technology components as well as policy, regulation, people, etc. The GILS is intended to help the public locate and access public information throughout the U.S. government.

The current GILS initiative builds upon a previous study, Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators (McClure, Ryan & Moen, 1992). That study, which was conducted for the Office of Management and Budget, the National Archives and Records Administration, and the General Services Administration, recommended that each agency establish a network-accessible locator that describes its information resources. The study also recommended that agencies use Z39.50 as the appropriate information retrieval protocol to achieve a distributed, standards-based Government Information Locator Service.

The development of the GILS Profile is documented in Using Z39.50 in an Application for the Government Information Locator Service (GILS) (McClure & Moen, 1994). The GILS Profile resulted from the work of a group comprising experts in Z39.50 implementations, system implementations, and information organization, and representatives of Federal agencies. The specifications included in the GILS Profile reflect the consensus of this group and input from a range of stakeholders.

3. Scope

The GILS Profile fully specifies the use of ANSI/NISO Z39.50 by the GILS. In addition, the GILS Profile provides the specifications for the overall GILS application relating to the GILS Core including other aspects of GILS conformant servers that are outside the scope of Z39.50.

This version of the GILS Profile focuses on requirements for a GILS server operating in the Internet environment. GILS clients will be able to interconnect with any GILS server, and these clients will behave in a manner that allows interoperability with the GILS server. Clients that support Z39.50 but do not implement the GILS Profile will be able to access GILS records with less than full GILS functionality.

The GILS Profile addresses many aspects of the GILS (e.g., intersystem interactions and information interchange) but does not specify user interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality.

4. Field of Application

The GILS Profile supports search and retrieval of GILS Locator Records contained in GILS servers by users in the Internet environment.

The GILS Profile will be used by developers of GILS servers. It will also be used by client developers to understand expected behaviors of GILS servers. A GILS server accessed using Z39.50 in the Internet environment acts primarily as a pointer to information resources. Some of these information resources pointed to by GILS Locator Records, as well as the GILS server itself, may be available electronically through other communications protocols including the common Internet protocols that facilitate electronic information transfer such as remote login (Telnet), File Transfer Protocol (FTP), and electronic mail (SMTP/MIME). The use of these protocols or other communications paths is outside the scope of the GILS Profile.

Once connected to a GILS server, users supported by appropriate clients that understand the GILS Profile may navigate through single or multiple servers. GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept a well-known search query and return a list of Locator Records in brief display format). Although the GILS Profile addresses GILS servers only, it is understood that clients have roles in the execution of these activities (e.g., browsing is also a client function in the sense of how it interprets and presents GILS data).

5. References

The following list contains documents that contain provisions which, through reference in this text, constitute provisions of the GILS Profile. At the time of this publication, the editions indicated were valid. All documents are subject to revision, and parties to agreements based on this Profile are warned against automatically applying any more recent editions of the documents listed below, since the nature of references made by the Profile to such documents, is that they may be specific to a particular edition. In addition, this list contains other documents that can be consulted for further information, background, etc.

[1] American National Standards Institute. (1985). American National Standard 39.2-1985 Bibliographic Information Interchange. New York: American National Standards Institute.

[2] Christian, Eliot. (1994, May 2). Government Information Locator Service (GILS): Report Information Infrastructure Task Force. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

[3] Lynch, Clifford A. (1994, April 30). "Using the Z39.50 Information Retrieval Protocol in the Internet Environment" [Draft RFC for Z39.50 over TCP/IP].

[4] McClure, Charles R. & Moen, William E. (1994, May 7). Using Z39.50 in an Application for the Government Information Locator Service (GILS). Available via anonymous FTP at <ericir.syr.edu> as /USGS/profile_background.doc.ps (Postscript format) and as /USGS/profile_background.doc.txt (ASCII text format).

[5] McClure, Charles R., Ryan, Joe & Moen, William E. Moen. (1992). Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators. 2 Vols. Bethesda, MD: National Audio Visual Center [Available from ERIC, document no. ED349031].

[6] National Information Standards Organization. (1992). ANSI/NISO Z39.50-1992, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. Gaithersburg, MD: NISO Press.

[7] National Institute of Standards and Technology. (1992). FIPS No. 173. Spatial Data Transfer Standard (August 28, 1992). Gaithersburg, MD: National Institute of Standards and Technology.

[8] Office of Management and Budget. (1993). Circular No. A-130, "Management of Federal Information Resources" (58 F.R. 36068, July 2,1993).

[9] Open Systems Environment Implementors Workshop/Special Interest Group on Library Applications (OIW/SIGLA). (1993). OIW/SIGLA Document #1: Using Z39.50-1992 Directly over TCP.

[10] RFC 1521, MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies.

[11] RFC 1522, MIME (Multipurpose Internet Mail Extensions) Part Two: Message Header Extensions for Non-ASCII Text.

[12] Uniform Resource Locators (URL): A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network. (1993, October). [Internet Draft]. The latest URL draft is: <url:ftp://info.cern.ch/pub/www/doc/url7a.txt>

[13] Uniform Resource Names. (1993, October). [Internet Draft]. The latest URN draft is: <url:ftp://ds.internic.net/internet-drafts/draft-ietf-uri-resource-names-01.txt>

[14] USMARC Format for Bibliographic Data. Washington, DC: Library of Congress, Cataloging Distribution Service.

6. Definitions

For purposes of this Profile, the following definitions apply.

Client: An initiating application. This application includes the Z39.50 origin.

Electronic Information Resource: Information resources that are maintained in electronic, digital format and may be accessed, searched, or retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM).

GILS Core: A subset of all GILS Locator Records which describe information resources maintained by the U.S. Federal government and comply with the defined GILS Core Elements and are mutually accessible through interconnected electronic network facilities without charge to the direct user.

Government Information: Information created, collected, processed, disseminated, or disposed of by or for the Federal government.

Government Information Locator Service (GILS) : A decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

Information Resource: Includes both government information and information technology.

Interoperability: A condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

BEST COPY AVAILABLE

Locator Record: A collection of related data elements describing an information resource, the information available in the resource, and how to obtain the information.

Mandatory: An element in a GILS Core Locator Record that must have a value provided by the record source. The GILS Profile does not specify which elements must be present from the perspective of GILS servers.

Origin: The part of a client application that initiates a Z39.50 association and is the source of requests during the association.

Profile: The statement of a function(s) and the environment within which it is used, in terms of a set of one or more standards, and where applicable, identification of chosen classes, subsets, options, and parameters of those standards. A set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

Registered Object: An object that is identified by a name-to-thing relationship in which the name is recorded by a registration authority to ensure that the names can be used unambiguously.

Server: An application that responds to an initiating application (i.e., a client). The application that includes the Z39.50 target.
Target: The part of an server application that accepts a Z39.50 association.

Uniform Resource Identifier (URI): A set of related standards for encoding resource location and identification information for electronic and other objects. Examples include Uniform Resource Locators (URLs) and Uniform Resource Names (URNs).

USMARC: An implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data. This documentation is published by the Library of Congress.

7. Z39.50 Specifications for GILS

This section details the required services available from Z39.50, describes an Attribute Set for searching, four Element Set Names by which the server presents some or all the elements (defined in the Schema) of the Locator Records, and prescribes the Record Syntaxes to be supported by GILS servers for the transfer of Locator Records.

7.1. Version

GILS clients and servers support Z39.50 Verson 2 as specified in Z39.50-1994. GILS requires support of various objects, some of which are not defined in Z39.50-1992. These are listed in 7.2.

7.2. GILS Objects

The following object identifier (OID) is assigned to the Z39.50 standard:

{iso (1) member-body (2) US (840) ANSI-standard-Z39.50 (10003)}.

This OID is abbreviated as: ANSI-standard-Z39.50.

Several object classes are assigned at the level immediately subordinate to ANSI-standard-Z39.50, including:

- 3 = attribute set definitions
- 4 = diagnostic definitions
- 5 = record syntax definitions
- 13 = database schema definitions.
- 14 = tagSet definitions.

GILS requires support of the following objects

- GILS attribute set: {ANSI-standard-Z39.50 3 3}
- bib1 diagnostic set: {ANSI-standard-Z39.50 4 1}
- USMARC record syntax: {ANSI-standard-Z39.50 5 10}
- SUTRS record syntax: {ANSI-standard-Z39.50 5 101}
- GRS-1 record syntax: {ANSI-standard-Z39.50 5 105}
- GILS schema: {ANSI-standard-Z39.50 13 2}
- tagSet-M {ANSI-standard-Z39.50 14 1}
- tagSet-G {ANSI-standard-Z39.50 14 2}.

7.3. Communication Services

When Transmission Control Protocol (TCP) is used as the transport service, the specification for use of TCP is found in OIW/SIGLA Document #1, "Using Z39.50-1992 Directly over TCP." The use of other communication services is not yet defined.

7.4. Z39.50 Services

There are three Z39.50 (Version 2) services that are required for conformance: Init, Search, and Present. No additional services are required for conformance to the GILS Profile. Other Z39.50 services, however, may be provided optionally by servers and used by clients.

Standard Z39.50 Init Service negotiation procedures control the use of all services.

7.4.1. Search

The GILS application will support Z39.50 Type 1 queries which are general purpose Boolean query structures.

7.4.1.1. Attribute Set

The GILS Attribute Set is a superset of the Bib-1 Attribute set and consists of all Bib-1 Attributes and additional Use Attributes that are defined for GILS elements (see Annex A for the GILS Use Attributes). These newly defined GILS Use Attributes are well-known and correspond semantically to GILS Core Elements. The GILS Attribute Set is a registered object.

GILS servers must support a limited number of GILS Attributes. The required GILS Attributes follow. (Note: The GILS Use Attribute is listed followed by the GILS Use Attribute Number and the corresponding GILS Core Element):

- Use Attributes: Local Number (12; Local Control Number); Author-name corporate (1005; Originator); Date/Time Last Modified (1012; Date of Last Modification); Record Source (1019; Record Source); Distributor Name (2001; Distributor Name); Index Terms -- Controlled (2002; Index Terms -- Controlled); Local Subject Index (29; Local Subject Term); Any (1016)

- Structure: Word (2), URx (104), Date (5), Word List (6)

- Relation: Greater than (5), Equal (3).

GILS servers should never return any of these four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes the combinations of required GILS Attributes listed in Table 1 in Annex A.

7.4.1.2. Well-known Search

To provide support for browsing GILS Locator Records, there is a well-known search consisting of the following GILS Attributes: Use Attribute: Local Number; Structure Attribute: URX; and a term of zero length. GILS servers that support browsing of records will create a result set of one or more GILS Locator Records that provide the necessary information to allow clients to offer menu-like displays of GILS Locator Records or other information and information resources.

The "Browse" in the GILS context involves only the Search and Present Services of Z39.50. "Browse" is used informally in the GILS Profile, and it is not related nor should it be confused with the Browse Facility or Scan Service of Z39.50.

7.4.2. Retrieval

This section describes the components and procedures used by Z39.50 to return records in response to a query.

7.4.2.1. Schema

The GILS Profile specifies a GILS Schema (see Annex D for the Schema). The GILS Schema is a registered object. The schema describes and/or defines tagSets used and an abstract record structure for a Locator Record. A schema in Z39.50 can be modified and may evolve over time, and it is reasonable to expect the GILS Schema will evolve.

The GILS Schema uses elements from tagSet-M and tagSet-G and defines in the GILS tagSet additional elements as necessary. The GILS Profile specifies tagTypes to identify tagSet-M elements (tagType = 1), tagSet-G elements (tagType =2), and the elements defined by the GILS tagSet (tagType = 4). Another tagType (tagType=3) is used to identify arbitrary string tags for locally defined elements.

The GILS tagSet element numbering begins with number 1. Elements can be nested and the tagging notation (i.e., the tag path) will reflect the nesting.

All well-known GILS Schema elements have assigned numeric tags. String-tags (i.e., text) may be used in the GILS Schema to label those elements that are not well-known (i.e., locally defined).

7.4.2.2. Element Sets Names

GILS servers will support four Element Set Names. GILS servers will interpret the use of the Element Set Names required by the GILS Profile to identify the following elements from the GILS Schema:

- The primitive element set name "B" contains at least: title, controlIdentifier, originator, and local control number
- The primitive element set name "G" contains: all B Element Set elements and crossReference
- The primitive element set name "W" contains: all B Element Set elements and bodyOfDisplay.
- The primitive element set name "F" contains: all elements available in the record.

The server should include in a retrieved record all of the elements specified by the element set name for which there is data available in the database record and which can be encoded in the requested record syntax (e.g., some types of locally defined binary data may not be encodable in a USMARC or SUTRS record).

7.4.2.3. Record Syntaxes

GILS servers are required to support the following three record syntaxes:

- USMARC -- an implementation of ANSI/NISO Z39.2 and maintained by the Library of Congress
- Generic Record Syntax (GRS-1) -- defined in Z39.50
- Simple Unstructured Text Record Syntax (SUTRS) -- defined in Z39.50.

Annex B contains a mapping of Core Elements to USMARC for use in the USMARC record syntax. However, since the data transformation is not fully reversible and requires interpretation, the record source is responsible for encoding the USMARC record(s).

The data in GILS Locator Records do not always map clearly into USMARC records, particularly when agencies add their own locally defined fields to the GILS Locator Record. This means that construction of USMARC records is subject to local interpretation. Therefore, GILS Locator Records in USMARC format obtained from other than the original record source should be considered non-definitive. The original source of the GILS Locator Record can be identified by examining the Original Control Identifier field of the record.

For interchange, GRS-1 records are to be treated as the complete and canonical representation; SUTRS and USMARC should be viewed as derivative records from the canonical representation and as such are not as complete or precise.

7.5. Preferred Display Format for Use with SUTRS

The GILS Profile recommends a preferred display format for SUTRS records (see Annex C for the recommended display format). For the SUTRS records, formatting instructions for a preferred display format is a concern of the server.
When the target transfers a GILS record using the SUTRS record syntax, it will encode the GILS record formatted according to the preferred display format, so that the client may present the record directly, without processing. For SUTRS, however, the client should not expect to be able to parse the record to obtain any individual GILS elements.

When the client presents a GILS record formatted by the server using the USMARC or GRS record syntax, it is recommended that the client consider the SUTRS suggested display layout in formatting the received record for presentation to the human end user.

7.6. Diagnostic Messages

The GILS application will use Diagnostic Set Bib-1.

8. Data Elements in the Locator Records

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The GILS Core Elements are defined in Annex E.

Annex A

GILS Attribute Set

The GILS Attribute Set is a superset of the Bib-1 Attribute Set and consists of all Bib-1 Attributes and the additional Use Attributes listed below. Additional Use Attributes that cannot be mapped to Bib-1 Use Attributes are numbered from 2000 through 2999. These are well-known Use Attributes.

GILS servers should never return any of these four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes the combinations of GILS Attributes listed in Table 1. An "X" in the table means that GILS servers will recognize and support this combination of Attributes.

USE	WORD	URx	DATE	WORD LIST	GREATER THAN	EQUAL
Local Number	X	X		X		X
Author-name corporate	X			X		X
Date/Time Last Modified			X		X	X
Record Source	X			X		X
Distributor Name	X			X		X
Index Term — Controlled	X			X		X
Local Subject Index	X			X		X
Any	X			X		X

TABLE 1
Recognized and Supported Combinations of GILS Attributes

As stated in 7.3.1.1, GILS servers are required to support a minimal set of Use Attributes. These are listed first. In the cases where a Bib-1 Use Attribute's Name is used, the corresponding GILS Core Element name appears in parentheses.

Required GILS Use Attributes

Use #	GILS Attribute Name
12	Local Number (Local Control Number)
29	Local Subject Index (Local Subject Term)
1005	Author-name corporate (Originator)
1012	Date/Time Last Modified (Date of Last Modification)
1016	Any
1019	Record Source
2001	Distributor Name
2002	Index Terms -- Controlled

Available GILS Use Attributes

Use #	GILS Attribute Name
4	Title
1007	Identifier - Standard (Control Identifier)
62	Abstract
2003	Purpose
2004	Access Constraints
2005	Use Constraints
2006	Distributor Organization
2007	Distributor Street Address
2008	Distributor City
2008	Distributor State
2010	Distributor Zip Code
2011	Distributor Country
2012	Distributor Network Address
2013	Distributor Hours of Service
2014	Distributor Telephone
2015	Distributor Fax
2016	Available Resource Description
2017	Available Order Process
2018	Available Technical Prerequisites
2019	Available Time Period -- Structured
2020	Available Time Period -- Textual
2021	Available Linkage
2022	Available Linkage Type
2023	Contact Name
2024	Contact Organization
2025	Contact Street Address
2026	Contact City
2027	Contact State
2028	Contact Zip Code
2029	Contact Country
2030	Contact Network Address
2031	Contact Hours of Service
2032	Contact Telephone
2033	Contact Fax
2034	Agency Program
2035	Sources of Data
2036	Thesaurus
2037	Methodology
2038	Bounding Rectangle -- Western-most

Available GILS Use Attributes

Use #	GILS Attribute Name
2039	Bounding Rectangle -- Eastern-most
2040	Bounding Rectangle -- Northern-most
2041	Bounding Rectangle -- Southern-most
2042	Geographic Keyword Name
2043	Geographic Keyword Type
2044	Time Period - Structured
2045	Time Period - Textual
2046	Cross Reference Title
2047	Cross Reference Linkage
2048	Cross Reference Type
2049	Original Control Identifier
2050	Supplemental Information

Annex B

GILS Core Element to USMARC Mapping

This Annex provides a mapping from GILS Core Elements to USMARC for use by the record source and GILS servers. Some of these data elements consist of two or more subelements, and this relationship is noted by the indentation.

Implementors should consult the authoritative documentation on USMARC found in USMARC Format for Bibliographic Data. The document is available from the Cataloging Distribution Service at the Library of Congress. A full description of the USMARC fields and available subfields within each field is in that document.

For some elements new USMARC fields and/or subfields may be incorporated into the USMARC format. New fields and/or subfields in the process of being considered for inclusion in USMARC are noted.

In cases where the 500 Note field is repeated to carry separate GILS Core Elements, the name of the GILS Core Element will be included and precede the data content for that field. A colon will separate the GILS Data Element name from the rest of the content in the field. For example, 500 Purpose: [data for this field]; 500 Agency Program: [data for this field]. Each such GILS Core Element should be carried in separate, repeating 500 fields.

In addition to the variable length fields listed in the mapping, a USMARC record will also include a Leader and field 008: Fixed-Length Data Elements. Certain character positions in each of these fixed length fields of a USMARC record will need to be coded specifically for GILS. In addition, USMARC records for GILS will include a code in the 042: Authentication Code to identify these USMARC records specifically as GILS Locator Records. The following suggest values for these fields (or parts of these fields):

Leader: A fixed field comprising the first 24 character positions (00-23) of each record that provides information for the processing of the record. For GILS records, the following character position is specifically relevant:

Character Position: 18 — Descriptive cataloging form
Value: # [i.e., blank] (Non-ISBD)
to indicate when International Standard Bibliographic Description is not followed.

008 Fixed Length Data Elements: Forty character positions (00-39) containing positionally-defined data elements that provide coded information about the record as a whole or about special bibliographic aspects of the item being cataloged. For GILS records that describe electronic information resources, the following character position is specifically relevant:

Character Position: 26 — Type of computer file

Values:
- a (Numeric data)
- b (Computer program)
- c (Representational)
- d (Document)
- e (Bibliographic data)
- f (Font)
- g (Game)
- h (Sound)
- i (Online system or service) [new code proposed]
- m (Combination)
- u (Unknown)
- z (Other)

042 Authentication Code

Value: gils [new code proposed]

GILS Data Elements and Corresponding USMARC Tags

GILS Data Element	USMARC Tag
Title	245$a
Control Identifier	001
Abstract	520
Purpose	500
Originator	710$a
Access Constraints	506
Use Constraints	540
Distributor	
Distributor Name	270$p [proposed field]
Distributor Organization	270$p [proposed field]
Distributor Street Address	270$a [proposed field]
Distributor City	270$b [proposed field]
Distributor State	270$c [proposed field]
Distributor Zip Code	270$e [proposed field]
Distributor Country	270$d [proposed field]
Distributor Network Address	270$m [proposed field]
Distributor Hours of Service	301$a [proposed field]
Distributor Telephone	270$k [proposed field]
Distributor Fax	270$l [proposed field]
Available Resource Description	037$f
Available Order Process	037$c
Available Technical Prerequisites	538
Available Time Period – Structured	045$c
Available Time Period – Textual	037$n [proposed field] (for non-electronic resource) 856$z (for electronic resource)

GILS Data Elements and Corresponding USMARC Tags

GILS Data Element	USMARC Tag
Available Linkage	856$u
Available Linkage Type	856 1st indicator/856$2
Point of Contact	856$m
	(for electronic resources)
Contact Name	270$p [proposed field]
Contact Organization	270$p [proposed field]
Contact Street Address	270$a [proposed field]
Contact City	270$b [proposed field]
Contact State	270$c [proposed field]
Contact Zip Code	270$e [proposed field]
Contact Country	270$d [proposed field]
Contact Network Address	270$m [proposed field]
Contact Hours of Service	301$a [proposed field]
Contact Telephone	270$k [proposed field]
Contact Fax	270$l [proposed field]
Record Source	040
Date Last modified	005
Agency Program	500
Sources of Data	537 [proposed field]
Index Terms -- Controlled	650
Thesaurus	650 1st indicator/ 650$2
Local Subject Term	653$a
Methodology	567
Spatial Reference	
Bounding Rectangle	255$c
Western-most	034$d
Eastern-most	034$e
Northern-most	034$f
Southern-most	034$g
Geographic Name	
Geographic Keyword Name	651
Geographic Keyword Type	655
Time Period -- Structured	045$c
Time Period -- Textual	513
Cross Reference Title	787$t
Cross Reference Linkage	787$w
Cross Reference Type	856 1st indicator/856$2
Original control identifier	035
Supplemental information	500

USMARC Tags and Field Names
(from USMARC Format for Bibliographic Data)

USMARC Tag	Subfield	Field Name
001		Control Number
005		Date and Time of Latest Transaction
034		Coded Cartographic Mathematical Data
	$d	Coordinates - westernmost longitude
	$e	Coordinates -- easternmost longitude
	$f	Coordinates -- northernmost latitude

USMARC Tags and Field Names
(from USMARC Format for Bibliographic Data)

USMARC Tag	Subfield	Field Name
	$g	Coordinates -- southernmost latitude
035		System Control Number
037		Source of Acquisition
	$b	Source of stock number/acquisition
	$c	Terms of availability
	$f	Form of issue
	$n	Note [proposed]
040		Cataloging Source
042		Authentication Code
245		Title Statement
	$a	Title
255		Cartographic Mathematical Data
	$c	Statement of coordinates
270	$a	Address
270	$b	City
270	$c	State or province
270	$d	Country
270	$e	Postal code
270	$k	Telephone number
270	$l	Fax number
270	$m	Electronic mail address
270	$p	Contact person
301	$a	Hours
500		General Note
506		Restrictions on Access Note
513		Type of Report and Period Covered Note
520		Summary, Etc. Note
537		Source of Data Note [proposed]
538		System Details Note
540		Terms Governing Use and Reproduction Note
567		Methodology Note
650		Subject Added Entry -- Topical Term
1st indicator		Level of subject
	$2	Source of heading or term
651		Subject Added Entry -- Geographic Name
653		Index Term -- Uncontrolled
	$a	Uncontrolled term
655		Index Term -- Genre/Form
710		Added Entry -- Corporate Name
	$a	Corporate name or jurisdiction name as entry element
787		Nonspecific Relationship Entry
	$t	Title
	$w	Record Control Number
856		Electronic Location and Access
1st indicator		Access method
	$m	Contact for access assistance
	$u	Uniform Resource Locator
	$z	Nonpublic note
	$2	Source of access

Annex C

Preferred Display Format for GILS Records

GILS servers will transfer records in three record syntaxes:

- USMARC
- Generic Record Syntax (GRS)
- Simple Unstructured Text Record Syntax (SUTRS).

In SUTRS, the formatting of the record contents is handled by the server, and the client receives a record devoid of structure. In USMARC and GRS, the record, whose structure is defined by the record syntax, is passed from the target to an orgin, and the client software has more flexibility in processing the record contents for display.

The recommended guidelines in this Annex describe how records should be displayed, whether formatted by the server or the client (but this does not preclude display formats in addition to the Preferred Display Format).

Record Organization:

The record should be organized so that the elements first viewed by the user provide adequate information to either choose or eliminate the record from further consideration. These elements are: Title, Originator, Controlled Vocabulary, Local Subject Index and Abstract.

Next in the order of presentation are elements that give detailed information about the information resource being described: Spatial Reference, Time Period, Availability, Sources of Data, Methodology, Access Constraints, Use Constraints, Point of Contact, and Supplemental Information.

The elements describing the reason for the existence of the data are next: Purpose and Agency Program.

Related information resources are listed next in the element: Cross Reference.

The final elements provide bibliographic control information: Control Identifier, Record Source, and Date of Last Modification.

General Instructions for Formatting Full Element Set Name Records:

All displayable elements are to be labelled with the full title of the field followed by a colon. Label mnemonics should only be used in situations where the user can ask for an explanation of the mnemonic. Mnemonics should not be used in SUTRS records, since it should be assumed that the client knows nothing about the server and is incapable of interpreting the mnemonics.

The subelements of constructed elements (i.e., locally defined fields, Availability, Spatial Reference, etc.) should be indented to reflect their association and structure within a well-structured element. Labels on subelements can eliminate the redundant leading parts (e.g., the word Available on the Availability subelements).

In the Controlled Vocabulary element, the Thesaurus subelement can be presented in parentheses, followed by the Index Terms. Multiple Index Terms should be separated by a semi-colon and a space (e.g., Controlled Vocabulary (MeSH): Kidney; Kidney Disease). Alternatively, the Thesaurus and Index Terms can be indented under the Controlled Vocabulary label, as is done with the other well-structured fields. Local Subject Terms should be separated by a semi-colon and a space.

Display Format for Brief Element Set Name Records:

Brief Records consist of the Title, Control Identifier, Originator, and Local Control Number fields. For display purposes, the Control Identifier and Local Control Number can be omitted. Brief Records may be formatted to fit on a single line. This may require that that one or both of the displayed fields will be truncated. Truncation can be indicated with with elipsis(...).

Display Format for G Element Set Name Records:

G Records consist of Brief Record elements and additionally, the Cross Reference element. For display purposes, the guidelines for Full Records should be followed.

Annex D

GILS Schema

The GILS Schema describes and defines tagSets and an Abstract Record Structure used with the Generic Record Syntax (GRS). The GILS Schema defines a GILS tagSet that associates a numeric tag with one or more GILS Core Elements.

Some GILS Core elements correspond to tags already defined in tagSet-M and tagSet-G, and these tags are used to identify GILS Core elements in the Abstract Record Structure. When the tagType is 1, the tag value is from tagSet-M. When the tagType is 2, the tag value is from tagSet-G. When the tagType is 3, the tag value is an arbitrary string tag. When the tagType is 4, the tag value is from the GILS tagSet.

There are two general classes of schema elements in the GILS Schema:

1) Primitive -- these elements cannot have locally defined subelements
2) Constructed -- these elements have one or more subelements any of which may be well-defined or target-defined; in the latter case, these locally defined subelements are identified with string tags

This Annex first presents the GILS tagSet that identifies the element, its unique tag, and a recommended datatype. This is followed by the GILS Abstract Record Structure that shows the full tag path for each element.

GILS tagSet

Tag	Element	Recommended Datatype
1	controlIdentifier	InternationalString
2	streetAddress	InternationalString
3	city	InternationalString
4	state	InternationalString
5	zipcode	InternationalString
6	hoursOfService	InternationalString
7	resourceDescription	InternationalString
8	technicalPrerequisites	InternationalString
9	westernMost	intUnit
10	easternMost	intUnit
11	northernMost	intUnit
12	southernMost	intUnit
13	geographicKeywordName	InternationalString
14	geographicKeywordType	InternationalString
15	timePeriodStructured	GeneralizedTime
16	timePeriodTextual	InternationalString

GILS tagSet

Tag	Element	Recommended Datatype
17	linkage	InternationalString
18	linkageType	InternationalString
19	recordSource	InternationalString
20	controlledTerm	InternationalString
21	thesaurus	InternationalString
22	localSubjectTerm	InternationalString
23	originalControlIdentifier	InternationalString

NOTE: The element "wellKnown" from tagSet-M (1,19) and referred to below has the following definition:

When an element is defined to be "structured into locally defined elements", the target may use this tag (i.e., wellKnown) in lieu of, or along with, locally defined tags. For example, an element named 'title' might be described to be "locally structured." The target might present the element structured into the following subelements: 'wellKnown', 'spineTitle', and 'variantTitle', where the latter two tags are target defined. In this case, 'wellKnown' is assumed to mean 'title'.

50	title	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
51	purpose	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
52	originator	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
53	accessConstraints	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
54	useConstraints	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
55	orderProcess	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
56	agencyProgram	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
57	sourcesOfData	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.

58	methodology	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
59	supplemental Information	Constructed as follows – This element may include the element wellKnown and may also include locally defined elements.
70	availability	Constructed as follows – This element may include any of the following as well as locally defined elements: distributor, resourceDescription, orderProcess, technicalPrerequisites, timePeriod, linkage, linkageType.
71	spatialReference	Constructed as follows – This element may include any of the following as well as locally defined elements: boundingRectangle, geographicName.
90	distributor	Constructed as follows – This element may include any of the following as well as locally defined elements: name, organization, streetAddress, city, state, zipCode, country, networkAddress, hoursOfService, phoneNumer, faxNumber.
91	boundingRectangle	Constructed as follows – This element may include any of the following as well as locally defined elements: westernMost, easternMost, northernMost, southernMost.
92	geographicName	Constructed as follows – This element may include any of the following as well as locally defined elements: geographicKeywordName, geographicKeywordType.
93	timePeriod	Constructed as follows – This element may include any of the following as well as locally defined elements: timePeriodStructured, timePeriodTextual.
94	pointOfContact	Constructed as follows – This element may include any of the following as well as locally defined elements: name, organization, streetAddress, city, state, zipCode, country, networkAddress, hoursOfService, phoneNumber, faxNumber.
95	controlled Vocabulary	Constructed as follows – This element may include any of the following as well as locally defined elements: indexTermsControlled, thesaurus.
96	indexTerms Controlled	Constructed as follows – This element may include the following as well as locally defined elements: controlledTerm.

97 localSubjectIndex Constructed as follows –

This element may include the following as well as locally defined elements: localSubjectTerm.

98 crossReference Constructed as follows –

This element may include any of the following as well as locally defined elements: title, linkage, linkageType.

GILS Abstract Record Structure

NOTE: The element "bodyOfDisplay" in tagSet-G (2,9) may be used by the target to combine into this single element (i.e., bodyOfDisplay) one or more of the elements from the following abstract record structure into a display format.

Tag path	Element	Mandatory?	Repeatable?
(1,10)	rank	N	N
(1,12)	url	N	N
(1,14)	local control number	Y	N
(1,16)	dateOfLastModification	Y	N
(4,50)	title	Y	N
(4,1)	controlIdentifier	Y	N
(2,6)	abstract	Y	N
(4,51)	purpose	Y	N
(4,52)	originator	Y	N
(4,53)	accessConstraints	Y	N
(4,54)	useConstraints	Y	N
(4,70)	availability	Y	Y
(4,70)/(4,90)	distributor	Y	N
(4,70)/(4,90)/(2,7)	distributorName	Y	N
(4,70)/(4,90)/(2,10)	distributorOrganization	Y	N
(4,70)/(4,90)/(4,2)	distributorStreetAddress	Y	N
(4,70)/(4,90)/(4,3)	distributorCity	Y	N
(4,70)/(4,90)/(4,4)	distributorState	Y	N
(4,70)/(4,90)/(4,5)	distributorZipCode	Y	N
(4,70)/(4,90)/(2,16)	distributorCountry	Y	N
(4,70)/(4,90)/(2,12)	distributorNetworkAddress	Y	Y
(4,70)/(4,90)/(4,6)	distributorHoursofService	Y	Y
(4,70)/(4,90)/(2,14)	distributorPhoneNumber	Y	Y
(4,70)/(4,90)/(2,15)	distributorFaxNumber	Y	Y
(4,70)/(4,7)	resourceDescription	N	N
(4,70)/(4,55)	orderProcess	Y	N
(4,70)/(4,8)	technicalPrerequisites	N	N
(4,70)/(4,93)	timePeriod	N	Y
(4,70)/(4,93)/(4,15)	timePeriodStructured	N	Y
(4,70)/(4,93)/(4,16)	timePeriodTextual	N	Y
(4,70)/(4,17)	linkage	N	N
(4,70)/(4,18)	linkageType	N	N
(4,94)	pointOfContact	Y	N
(4,94)/(2,7)	contactName	Y	N
(4,94)/(2,10)	contactOrganization	Y	N

Tag path	Element	Mandatory?	Repeatable?
(4,94)/(4,2)	contactStreetAddress	Y	N
(4,94)/(4,3)	contactCity	Y	N
(4,94)/(4,4)	contactState	Y	N
(4,94)/(4,5)	contactZipCode	Y	N
(4,94)/(2,16)	contactCountry	Y	N
(4,94)/(2,12)	contactNetworkAddress	Y	Y
(4,94)/(4,6)	contactHoursofService	Y	Y
(4,94)/(2,14)	contactPhoneNumber	Y	Y
(4,94)/(2,15)	contactFaxNumber	Y	Y
(4,19)	recordSource	Y	N
(4,56)	agencyProgram	N	N
(4,57)	sourcesOfData	N	N
(4,95)	controlledVocabulary	N	Y
(4,95)/(4,96)	indexTermsControlled	Y	N
(4,95)/(4,96)/(4,20)	controlledTerm	Y	Y
(4,95)/(4,21)	thesaurus	Y	N
(4,97)	localSubjectIndex	N	N
(4,97)/(4,22)	localSubjectTerm	Y	Y
(4,58)	methodology	N	N
(4,71)	spatialReference	N	N
(4,71)/(4,91)	boundingRectangle	N	N
(4,71)/(4,91)/(4,9)	westernMost	N	N
(4,71)/(4,91)/(4,10)	easternMost	N	N
(4,71)/(4,91)/(4,11)	northernMost	N	N
(4,71)/(4,91)/(4,12)	southernMost	N	N
(4,71)/(4,92)	geographicName	N	Y
(4,71)/(4,92)/(4,13)	geographicKeywordName	Y	N
(4,71)/(4,92)/(4,14)	geographicKeywordType	Y	N
(4,93)	timePeriod	N	Y
(4,93)/(4,15	timePeriodStructured	N	N
(4,93)/(4,16)	timePeriodTextual	N	N
(4,98)	crossReference	N	Y
(4,98)/(4,50)	crossReferenceTitle	Y	N
(4,98)/(4,17)	crossReferenceLinkage	Y	N
(4,98)/(4,18)	crossReferenceType	Y	N
(4,23)	originalControlIdentifier	Y	N
(4,59)	supplementalInformation	Y	N

Annex E

GILS Core Elements

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The term "mandatory" as used in this Profile applies to administraton of the subset of GILS Locator Records that have been identified by the record source as participating in the GILS Core. GILS servers are not required to distinguish "mandatory" from other elements.

TITLE (Mandatory, Not Repeatable): This element conveys the most significant aspects of the referenced resource and is intended for initial presentation to users independently of other elements. It should provide sufficient information to allow users to make an initial decision on likely relevance. It should convey the most significant information available, including the general topic area, as well as a specific reference to the subject.

CONTROL IDENTIFIER (Mandatory, Not Repeatable): This element is defined by the information provider and is used to distinguish this locator record from all other GILS Core locator records. The control identifier should be distinguished with the record source agency acronym as provided in the U.S. Government Manual.

ABSTRACT (Mandatory, Not Repeatable): This element presents a narrative description of the information resource. This narrative should provide enough general information to allow the user to determine if the information resource has sufficient potential to warrant contacting the provider for further information. The abstract should not exceed 500 words in length.

PURPOSE (Mandatory, Not Repeatable): This element describes why the information resource is offered and identifies other programs, projects, and legislative actions wholly or partially responsible for the establishment or continued delivery of this information resource. It may include the origin and lineage of the information resource, and related information resources.

ORIGINATOR (Mandatory, Not Repeatable): This element identifies the information resource originator, named as in the U.S. Government Manual where applicable.

ACCESS CONSTRAINTS (Mandatory, Not Repeatable): This element in some cases may contain the value "None." It describes any constraints or legal prerequisites for accessing the information resource or its component products or services. This includes any access constraints applied to assure the protection of privacy or intellectual property, and any other special restrictions or limitations on obtaining the information resource. Guidance on obtaining any users' manuals or other aids needed for the public to reasonably access the information resource must also be included here.

USE CONSTRAINTS (Mandatory, Not Repeatable): This element in some cases may contain the value "None." It describes any constraints or legal prerequisites for using the information resource or its component products or services. This includes any use constraints applied to assure the protection of privacy or intellectual property and any other special restrictions or limitations on using the information resource.

AVAILABILITY (Mandatory, Repeatable): This element is a grouping of subelements that together describe how the information resource is made available.

DISTRIBUTOR (Mandatory, Not Repeatable): This subelement consists of the following subordinate fields that provide information about the distributor:

- **DISTRIBUTOR NAME**
- **DISTRIBUTOR ORGANIZATION**
- **DISTRIBUTOR STREET ADDRESS**
- **DISTRIBUTOR CITY**
- **DISTRIBUTOR STATE**
- **DISTRIBUTOR ZIP CODE**
- **DISTRIBUTOR COUNTRY**
- **DISTRIBUTOR NETWORK ADDRESS**
- **DISTRIBUTOR HOURS OF SERVICE**
- **DISTRIBUTOR TELEPHONE**
- **DISTRIBUTOR FAX**

RESOURCE DESCRIPTION (Optional, Not Repeatable): This subelement identifies the resource as it is known to the distributor.

ORDER PROCESS (Mandatory, Not Repeatable): This subelement provides information on how to obtain the information resource from this distributor, including any fees associated with acquisition of the product or use of the service, order options (e.g., available in print or digital forms, PC or Macintosh versions), order methods, payment alternatives, and delivery methods.

TECHNICAL PREREQUISITES (Optional, Not Repeatable): This subelement describes any technical prerequisites for use of the information resource as made available by this distributor.

AVAILABLE TIME PERIOD (Optional, Repeatable): This subelement provides the time period reference for the information resource as made available by this distributor, in one of two forms:

TIME PERIOD -- STRUCTURED: Time described using the USMARC prescribed structure.

TIME PERIOD -- TEXTUAL: Time described textually.

AVAILABLE LINKAGE (Optional, Not Repeatable): This subelement provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information.

AVAILABLE LINKAGE TYPE (Optional, Not Repeatable): This subelement occurs if there is an Available Linkage described. It provides the data content type (i.e., MIME) for the referenced URI.

POINT OF CONTACT FOR FURTHER INFORMATION (Mandatory, Not Repeatable): This element identifies an organization, and a person where appropriate, serving as the point of contact plus methods that may be used to make contact. This element consists of the following subelements:

CONTACT NAME
CONTACT ORGANIZATION
CONTACT STREET ADDRESS
CONTACT CITY
CONTACT STATE
CONTACT ZIP CODE
CONTACT COUNTRY
CONTACT NETWORK ADDRESS
CONTACT HOURS OF SERVICE
CONTACT TELEPHONE
CONTACT FAX.

RECORD SOURCE (Mandatory, Not Repeatable): This element identifies the organization, as named in the U.S. Government Manual, that created or last modified this locator record.

DATE OF LAST MODIFICATION (Mandatory, Not Repeatable): This element identifies the latest date on which this locator record was created or modified.

AGENCY PROGRAM (*, Not Repeatable): This element identifies the major agency program or mission supported by the system and should include a citation for any specific legislative authorities associated with this information resource.

* This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system.

SOURCES OF DATA (*, Not Repeatable): This element identifies the primary sources or providers of data to the system, whether within or outside the agency.

* This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system.

CONTROLLED VOCABULARY (Optional, Repeatable): This element is a grouping of subelements that together provide any controlled vocabulary used to describe the resource and the source of that controlled vocabulary:

INDEX TERMS -- CONTROLLED (Optional, Not Repeatable): This subelement is a grouping of descriptive terms drawn from a controlled vocabulary source to aid users in locating entries of potential interest. Each term is provided in the subordinate repeating field:

CONTROLLED TERM.

THESAURUS (Optional, Not Repeatable): This subelement provides the reference to a formally registered thesaurus or similar authoritative source of the controlled index terms. Notes on how to obtain electronic

access to or copies of the referenced source should be provided, possibly through a Cross Reference to another locator record that more fully describes the standard and its potential application to locating GILS information.

LOCAL SUBJECT INDEX (Optional, Not Repeatable): This element is a grouping of descriptive terms to aid users in locating resources of potential interest, but the terms are not drawn from a formally registered controlled vocabulary source. Each term is provided in the repeating subelement:

LOCAL SUBJECT TERM

METHODOLOGY (Optional, Not Repeatable): This element identifies any specialized tools, techniques, or methodology used to produce this information resource. The validity, degree of reliability, and any known possibility of errors should also be described.

SPATIAL REFERENCE (Optional, Not Repeatable): This element is a grouping of subelements that together provide the geographic reference for the information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."

BOUNDING RECTANGLE (Optional, Not Repeatable): This subelement provides the limits of coverage expressed by latitude and longitude values in the order:

WESTERN-MOST
EASTERN-MOST
NORTHERN-MOST
SOUTHERN-MOST.

GEOGRAPHIC NAME (Optional, Repeatable): This subelement identifies significant areas and/or places within the coverage through two associated constructs:

GEOGRAPHIC KEYWORD NAME
GEOGRAPHIC KEYWORD TYPE.

TIME PERIOD OF CONTENT (Optional, Repeatable): This element provides time frames associated with the information resource, in one of two forms:

TIME PERIOD – STRUCTURED: Time described using the USMARC prescribed structure.
TIME PERIOD – TEXTUAL: Time described textually.

CROSS REFERENCE (Optional, Repeatable): This element is a grouping of subelements that together identify another locator record likely to be of interest:

CROSS REFERENCE TITLE (Mandatory, Not Repeatable): This subelement provides a human readable textual description of the cross reference.

CROSS REFERENCE LINKAGE (Mandatory, Not Repeatable): This subelement provides the machine readable information needed to perform the access (i.e., URI).

CROSS REFERENCE TYPE (Mandatory, Not Repeatable): This subelement occurs if there is a CROSS REFERENCE LINKAGE and0 provides the data content type (i.e., MIME) for the referenced URI.

ORIGINAL CONTROL IDENTIFIER (Optional, Not Repeatable): This element is used by the record source to refer to another GILS locator record from which this locator record was derived.

SUPPLEMENTAL INFORMATION (Optional, Not Repeatable): Through this element, the record source may associate other descriptive information with the GILS Core locator record.

Attachment C

USING Z39.50 IN AN APPLICATION FOR THE GOVERNMENT INFORMATION LOCATOR SERVICE (GILS): A BACKGROUND PAPER

Developed as Part of the Cooperative Research Study:

Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard

between the
School of Information Studies, Syracuse University
and
The United States Geological Survey
funded by
The Interagency Working Group on Data Management for Global Change

Charles R. McClure <cmcclure@suvm.acs.syr.edu>
William E. Moen<wemoen@mailbox.syr.edu>
Co-Principal Investigators

School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244-4100
Telephone: (315) 443-2911
Fax: (315) 443-5806

May 7, 1994

CONTENTS

1. Introduction

2. ANSI/NISO Z39.50: A Standard for Information Retrieval

3. The Research and Development Project

   3.1. A Profile for GILS

4. Assumptions and Agreements about GILS

   4.1. The GILS System Architecture Model
   4.2. Navigating through GILS
   4.3. Uniform Resource Identifiers in GILS
   4.4. GILS Locator Records
   4.5. GILS Searching
   4.6. GILS Browsing
   4.7. Input Formats for GILS Records
   4.8. Use of Z39.50 in GILS

5. Z39.50 SPECIFICATIONS FOR THE GILS APPLICATION

   5.1. Version
   5.2. GILS Objects
   5.3. Communications Services
   5.4. Z39.50 Facilities and Services
   5.4.1. Search
   5.4.1.1. Attribute Set
   5.4.1.2. Well-known Search
   5.4.2. Retrieval
   5.4.2.1. Schema
   5.4.2.2. Element Set Names
   5.4.2.3. Record Syntaxes
   5.5. Preferred Display Format for Use with SUTRS
   5.6. Diagnostic Messages

6. Data Elements in GILS Locator Records

7. Conclusion

Notes

References

Appendices

   Appendix A — Project Team Members
   Appendix B — Definitions

USING Z39.50 IN AN APPLICATION FOR THE GOVERNMENT INFORMATION LOCATOR SERVICE (GILS)

1. Introduction

This document describes a research effort focused on the use of ANSI/NISO Z39.50, the American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (National Information Standards Organization, 1992), in the proposed Government Information Locator Service (GILS). A primary component of this research has been the development of a GILS Profile.

The GILS is a response to the need for users to be able to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources. The authoritative document describing the vision and function of GILS is Government Information Locator Service (GILS) (Christian, 1994); that document provides an overview of GILS, its objectives, service requirements, and core requirements.[1] According to Christian (1994) the GILS is a decentralized collection of locators and associated information services that includes information and technology components as well as policy, regulation, and people. The GILS is intended to help the public locate and access public information throughout the U.S. government.

The GILS Profile includes the specifications for Z39.50 in the GILS application operating in the Internet environment.[2] Additionally, the GILS Profile addresses other aspects of GILS conformant servers that are beyond the scope of Z39.50. The GILS Profile provides the specifications for the overall GILS application relating to the GILS Core, which is a subset of all GILS Locator Records, and completely specifies the use of Z39.50 in this application.[3] The GILS Profile will be used by implementations of GILS servers. It will also be used by client developers to understand expected behaviors of GILS servers.

This paper discusses the work by project team of Z39.50 experts and other participants in developing the GILS Profile. Components of the team's work included understanding the high-level functional requirements for GILS described in Christian (1994), agreeing upon a model of the GILS system architecture and information flows in the GILS, and delineating the functional requirements that could be addressed by the GILS Profile. This paper is intended to serve as background to the assumptions, choices, and decisions by the project team that resulted in the specifications contained in the GILS Profile.

The work of the project team occurred within the research project, "Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard," coordinated by Syracuse University and the United States Geological Survey, funded by the Interagency Working Group on Data Management for Global Change. The result of this research project is to provide the specifications (i.e., the GILS Profile) for initial GILS implementations that are expected to provide users with information about the location of and ways to access or acquire Federal information resources.

The current research builds upon a previous study, Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators (McClure, Ryan & Moen, 1992). That study, which was conducted for the Office of Management and Budget, the National Archives and Records Administration, and the General Services Administration, recommended that each Federal agency establish a network-accessible locator that describes its information resources. The study also recommended that agencies use Z39.50 as the appropriate information retrieval protocol to achieve a distributed, standards-based Government Information Locator Service.

2. ANSI/NISO Z39.50: A Standard for Information Retrieval

The information retrieval protocol, Z39.50, provides a common language for clients and servers to select and retrieve records from databases. The purpose of Z39.50 is to allow one computer operating in a client mode to perform information retrieval queries against another computer acting as an information server; the protocol

also provides for the transfer of records or other information from the server to the client. Z39.50 does not prescribe how a particular system will execute the searching and retrieval on databases nor does it prescribe user interface requirements.

The standard is an applications-layer protocol within the Open Systems Interconnection (OSI) reference model. The OSI Basic Reference Model (ISO 7498: 1984 Information processing systems—Open Systems Interconnection--Basic reference model) was developed at the international level by the International Organization for Standardization (ISO).

ANSI/NISO Z39.50 is an American National Standard developed and was approved by the National Information Standards Organization (NISO) in 1988. NISO balloted and approved a 1992 revision of the standard (also referred to as Version 2). The Z39.50 Implementors Group (ZIG) has been preparing Version 3 of the standard (which contains new enhancements, extensions, etc.) for official balloting through NISO in 1994.[4]

Z39.50 is parallel to two OSI international standards: ISO 10162: 1993 Information and documentation--Search and Retrieve Application Service Definition; and ISO 10163-1: 1993 Information and documentation--Search and Retrieve Application Protocol Specification.

Although developed as an OSI application-layer protocol, Z39.50 is currently used by implementors in the Transmission Control Protocol (TCP) environment of the Internet. The success of an Z39.50 interoperability testbed in 1992 showed how the transport service of TCP can successfully support the protocol. Lynch (1994) provides the specification for using Z39.50 over TCP.

3. The Research and Development Project

For the current study, a group comprising experts in Z39.50 implementations, system implementations, and information organization, and representatives of Federal agencies has been working as part a research project coordinated by Syracuse University and the United States Geological Survey, funded by the Interagency Working Group on Data Management for Global Change. (See Appendix A for names of project team members.) To advance the development of the GILS, the research project has focused on the use of open systems standards to improve the utility of information searching and retrieval on digital networks. More specifically, the project has as its objectives to:

- Expand research and development on the American National Standard for information searching and retrieval (Z39.50) for its application in facilitating public access to Federal information resources and speeding the development of interoperable systems
- Build consensus of major stakeholders on the manner in which Z39.50 can be applied in GILS implementations
- Develop an application profile for networked-based GILS implementations that references Z39.50 and other relevant standards for use in the Internet environment
- Support and encourage test implementations of the profile by interested parties to provide evaluations of the profile and for interoperability testing.

To achieve these objectives, the project team focused it primary attention on developing the GILS Profile.

3.1. A Profile for GILS

A profile is "a set of one or more base standards, and where applicable, the identification of chosen classes, subsets, options and parameters of those base standards, necessary for accomplishing a particular function

(International Organization for Standardization/International Electrotechnical Commission, 1992, p. 2). Profiles are also referred to as "functional standards," "implementation agreements," or "specifications." Since open systems standards often include choices and options, profiles specify the values and parameters of a standard for an application or implementation to increase the likelihood of interoperability and interworking. A profile, then, is a set of implementation agreements that guide implementors in applying one or more standards in a specific and limited context.

The research team broadened this definition for the GILS Profile to include not only the specifications for Z39.50 and other relevant standards in the application but also other aspects of a GILS conformant server that are beyond the scope of these standards. The GILS Profile does provide the specifications for the overall GILS application relating to the GILS Core and completely specifies the use of Z39.50 in this application. The GILS Profile will facilitate interoperability of independently developed components of the GILS Core. Further, in developing the GILS Profile, the project team was aware of the need to understand and address interoperability issues with the currently installed base of available implementation technology.

This first version of the GILS Profile focuses on the requirements for a GILS server operating in the Internet environment. GILS clients will be able to interconnect with any GILS server, and these clients will behave in a manner that allows interoperability with the GILS server. Clients that support Z39.50 but do not implement the GILS Profile should be able to access GILS records but with less than full GILS functionality. Although the GILS Profile addresses GILS servers only, it is understood that clients have roles in the execution of information retrieval activities.

The GILS Profile addresses many aspects of the GILS (e.g., intersystem interactions and information interchange) but does not specify user interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality. These aspects are also outside the scope of Z39.50.

The Government Information Locator Service (GILS) (Christian, 1993) (hereafter referred to as the GILS document) provided the research team with high-level requirements for the GILS. Based on those requirements, the research team delineated assumptions about the operation and information flows of the GILS and developed functional requirements for the GILS. This process allowed the research team to identify a subset of Z39.50 and other existing and emerging standards that would support these functional requirements. The following sections of this document detail the research team's assumptions, model, conclusions, and Z39.50 specifications.

4. Assumptions and Agreements about GILS

The GILS document presents an overview of GILS, including its objectives, service requirements, and core requirements.[5] These requirements, however, are often described in general terms rather than in terms of specific functional requirements. The research team proceeded to develop an interpretation and understanding of the high-level requirements presented in the GILS document. As a result, the team delineated the functional requirements that could be addressed by the GILS Profile. To accomplish this, the research team agreed upon a model of the system architecture that adequately described the GILS operation and information flows. In addition, the research team also reached other consensus agreements on the use of Z39.50 and other existing or emerging standards (e.g., USMARC, standards such as the Uniform Resource Identifiers developed for the Internet environment by the Internet Engineering Task Force [IETF], etc.).[6]

4.1. The GILS System Architecture Model

The GILS is understood to be an agency-based, Internet-accessible locator service. "Direct users" (see GILS document, p. 4) will connect to GILS servers via the Internet to find information about a wide range of Federal information resources. Once connected to a GILS server, users supported by appropriate clients that understand the GILS Profile, may navigate through single or multiple servers. GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept

a well-known search query and return a list of Locator Records in brief display format). The use of the national standard for network information retrieval, Z39.50, provides for interoperability between clients and multiple servers.

Agencies will develop and maintain GILS servers. These GILS servers are machine-readable databases that contain Locator Records describing Federal information resources. These decentralized agency-based GILS servers enable ongoing maintenance responsibilities to be carried out by those who understand and manage the information resources. The GILS, then, is a distributed resource consisting of agency-based servers. The GILS Profile does not specify the base technology (e.g., a database management system) that an agency uses to mount its database of Locator Records nor does it specify internal storage structures for Locator Records in the database.

According to McClure, Ryan & Moen (1992, p. 2), a locator is a "machine-readable database that identifies different information resources (e.g., databases, libraries, clearinghouses, print publications, bulletin boards, etc.) and describes the information available in these resources. Usually, the locator does not provide the actual information, but rather points the user to the information sources that do provide the needed information." The GILS document states that "GILS is an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information" (p. 4).

A GILS server accessed using Z39.50 in the Internet environment acts primarily as a pointer to information resources. The GILS server, as well as some of the information resources pointed to by GILS Locator Records, may be available electronically through other communications protocols including the common Internet protocols that facilitate electronic information transfer such as remote login (Telnet), File Transfer Protocol (FTP), and electronic mail (SMTP/MIME). The use of these protocols or other communications paths is outside the scope of this project and of the GILS Profile.

The public will use the GILS either directly or through intermediaries (the intermediaries obtain GILS information as direct users themselves or from other intermediaries). The GILS document (p. 4) describes these two classes of users. The concern for the project team, however, was limited to "direct users" accessing the GILS via the Internet using client/server implementations that rely upon Z39.50 as the information retrieval protocol.

GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept a well-known search query and return a list of Locator Records in brief display format). Although the GILS Profile addresses GILS servers only, it is understood that clients have roles in the execution of these activities (e.g., browsing is also a client function in the sense of how it interprets and presents GILS data). The server should include in a retrieved record all elements or combinations of elements of the database record for which there is data available and which can be encoded in the requested record syntax (see Sections 5.4.2.2. — Element Set Names and 5.4.2.3. — Record Syntax).

4.2. Navigating through GILS

Direct users must have prior knowledge of at least one GILS server and its network address, and must be able to access it to enter the GILS. Upon entry, however, users supported by appropriate clients that understand the GILS Profile may navigate through single or multiple GILS servers by following the links provided in the Locator Records (see Section 4.4.).

The semantics of the Locator Records coupled with a client that understands these semantics and building upon the ability of the Z39.50 protocol to provide a uniform interface to multiple autonomously managed servers combine to provide the user with the impression of seamless navigation among these distributed servers. The semantics of the Locator Records facilitate elimination of duplicate records, further fostering the impression of a single system built out of autonomous, distributed servers.

Each GILS server can be represented by a Locator Record in other GILS servers. Some of these servers will include references to all other GILS servers, and these might be regarded as a kind of "directory of directories." However, GILS itself does not assign any hierarchical status to specific servers nor does it specify a "root server." Rather, the structure and content of the GILS Locator Records enable, for example, the aggregation of Locator Records in "directories" that could be offered by one or more Federal agencies or other organizations.

4.3. Uniform Resource Identifiers in GILS

GILS incorporates the use of Uniform Resource Identifiers (URIs) to improve interoperability and navigation in the Internet environment. URIs comprise a set of related standards for encoding resource location and identification information for electronic and other objects. The URI Working Group of the IETF defines and specifies URIs.[5] There are currently three objects within the URI set: the Uniform Resource Locator (URL) (1993); the Uniform Resource Name (URN) (1993); and Uniform Resource Characteristics (URC). The URI Working Group has approved URLs for experimental standardization, and it is expected to approve URNs in 1994. URCs are in the developmental stages.

GILS Locator Records contain fields for URIs. A scenario for the GILS as specifically related to URIs would be: A user, via a client, browses or searches a set of GILS servers and is presented with a set of GILS Locator Records, each referring to information resources (including other GILS servers) or related GILS Locator Records. As the user reads through the records, embedded URIs provide the ability for the client to directly access these described resources, related Locator Records, or GILS servers. URIs can serve as a direct reference to related works (e.g., a cross reference to another resource).

By incorporating the use of URIs, the GILS is facilitating interoperability within the wider Internet community while accomplishing its goal of providing improved access to Federal information resources.

4.4. GILS Locator Records

A GILS server contains individual Locator Records; these well-structured Locator Records include a standardized set of data elements (see Section 6 — Data Elements in GILS Locator Records). The data elements provide summary descriptions of Federal information resources. GILS servers (i.e., machine-readable databases) are themselves Federal information resources and can be described by Locator Records.

Locator Records in a single agency's (e.g., Agency A) GILS server can represent one of the following:

1) An internal information resource of Agency A. The primary purpose of the GILS server at Agency A is to provide Locator Records describing its own information resources. Agency A's GILS server is an information resource of Agency A, so Agency A's server may contain a Locator Record describing this GILS server.

2) Any information resource external to Agency A. This includes information resources (including another agency's GILS server) that are described in Locator Records by other agencies participating in GILS or any other information resources Agency A's GILS server providers wish to describe.

The distributed design of the GILS is partly supported by records in case #2. These records may provide specific links between GILS servers.

A Locator Record consists of a number of data elements that identify and describe an information resource. (Core Elements are noted in uppercase letters throughout this document.) Several data elements can be included in Locator Records to facilitate GILS navigation and network-based access to information:

- Each retrieved Locator Record contains a LOCAL CONTROL NUMBER generated by the system and guaranteed to be unique on the server from which the Locator Record is retrieved.

- Each Locator Record contains a CONTROL IDENTIFIER in the form of a Uniform Resource Identifier (URI). Agency A's server may contain Locator Records with CONTROL IDENTIFIERS that identify Locator Records from other Agencies' servers. This data element allows GILS Locator Records to be replicated on multiple servers for the convenience of GILS users.

- Each Locator Record contains an AVAILABILITY element that informs the user how to procure the described information resource. If the information resource is an electronic information system or electronic document, the AVAILABILITY element includes AVAILABLE LINKAGE information in both human- and machine-readable form. The network linkage information may be used to connect to and access the electronic information resource.

Different agencies may create or offer Locator Records describing the same information resource (these may be existing Locator Records that have been replicated and/or modified, or entirely new Locator Records). These multiple records can offer different views of a single resource from the particular perspectives of the agencies creating/modifying a Locator Record. For example, two agencies may wish to highlight different aspects of the content of a specific information resource and to describe it in terms common to an agency's particular user community. Each agency will assign its own CONTROL IDENTIFIER to the Locator Record it creates or substantially modifies.

An agency (Agency B) may copy another agency's (Agency A) Locator Record. These are considered replicated records. In this case, two things might happen:

1) If Agency B makes no substantive changes to the replicated Locator Record from Agency A, the CONTROL IDENTIFIER is not changed.

2) If Agency B makes substantive changes to the replicated Locator Record from Agency A , a new CONTROL IDENTIFIER is assigned by the agency (Agency B) making the change. The CONTROL IDENTIFIER assigned by Agency A is retained in Agency B's new record in the data element ORIGINAL CONTROL IDENTIFIER.

This process of replication and modification may become very complex, and the inclusion of the ORIGINAL CONTROL IDENTIFIER is intended to enable the user to trace the location of the record created by the original source of the information resource.

4.5. GILS Searching

Users will be able to search a GILS server as a means of finding out how to acquire or access the information resource described by one or more Locator Records. GILS servers may support a variety of search strategies including those:

- to find known items (e.g., where the user knows the exact TITLE of an information resource described in a Locator Record)

- to find resources whose Locator Records contain certain words or phrases

- to find resources by topic (e.g., using a controlled vocabulary)

- to find resources whose Locator Records meet other criteria (e.g., specific ORIGINATOR agencies).

A user's search specification is received by a GILS server using the Search Facilities of Z39.50. The searchable elements of the Locator Records correspond to Attributes (described in Section 5.4.1.1. — Attribute Set). The exact manner by which the user constructs the query is an interface issue and not specified by the GILS Profile, but users supported by appropriate clients that understand the GILS Profile should be able to specify searches with each of the required Attributes listed in Section 5.4.1.1.

As a GILS server completes a search, it produces a result set and makes that available to a client. The GILS server provides the client the contents of selected records from the result set using the Present Service of Z39.50. The GILS server must respond to requests that records be presented in any of three Record Syntaxes (see Section 5.4.2.3. — Record Syntaxes) mandated by the GILS Profile and one of the four Element Set Names (see Section 5.4.2.2. — Element Set Names) specified by the GILS Profile. The exact manner in which records are presented to the user is an interface issue and not within the scope of the GILS Profile.

4.6. GILS Browsing

A GILS server may provide a structure for browsing that is comprised of a chain of Locator Records traversed through pointers specified in the GILS Core Element CROSS REFERENCE. The CROSS REFERENCE is a repeating element. Each occurrence contains a item pointer in the form of a Uniform Resource Identifier (URI), the title of the item, and a content type to identify it. Each referenced item may be a Locator Record on the same GILS server or on another GILS server.

To provide support for browsing GILS Locator Records, there is a well-known search consisting of specific GILS Attributes and a term of zero length. GILS servers that support browsing of records will create a result set of one or more GILS Locator Records that provide the necessary information to allow clients to offer menu-like displays of GILS Locator Records or other information and information resources.

The well-known search allows users to browse a GILS server when or if they have no other starting point. If a particular GILS server does not support browsing, the response to the well-known search may be an error message or an empty result set (i.e., this particular server does not contain any such records that match the query requirements).

4.7. Input Formats for GILS Records

The GILS Profile does not recommend or prescribe any formats for records input to the software that feeds a particular GILS server database. This is a concern for GILS application developers, those who create the records, and/or those who load the records from other existing systems.

4.8. The Use of Z39.50 in GILS

Z39.50 provides a key part of the foundation for the GILS. This standard enables the interoperability of a variety of systems and hardware platforms in a client/server environment for the purposes of information retrieval. The GILS Profile will include the complete specifications of a subset of Z39.50 for use in the GILS application. The GILS Profile, in addition, will specify necessary characteristics of the GILS application that are outside the scope of Z39.50 including reference to other existing or emerging standards. Separate implementations will have an improved likelihood of interoperability and interworking when they conform to a common profile.

5. Z39.50 Specifications for the GILS Application

Based on the descriptions of the GILS system architecture model outlined above, the project team determined how Z39.50 will support the functional requirements of GILS. The specifications for using Z39.50 is documented in the GILS Profile. The GILS Profile is the authoritative source for Z39.50 specification for the GILS application and should be referred to for completeness and accuracy of specifications.

The GILS Profile details the required facilities and services available from Z39.50, describes an Attribute Set for searching Locator Records and four Element Sets by which the server presents some or all the elements of the Locator Records, and prescribes three Record Syntaxes to be supported by GILS servers for the transfer of Locator Records. This section outlines the Z39.50 specifications for the GILS Profile.

The terminology and concepts presented in this section are specific to Z39.50. Readers should consult the complete standard (National Information Standards Organization, 1992) for further information and reference. For example, the standard uses the words "origin" and "target," rather than "client" and "server."

5.1. Version

GILS origin (clients) and targets (servers) support Z39.50 Version 2 as specified in Z39.50-1994. GILS requires support of various objects, some of which are not defined in Z39.50-1992. These are listed in 7.2.

5.2. GILS Objects

The following object identifier (OID) is assigned to the Z39.50 standard:

{iso (1) member-body (2) US (840) ANSI-standard-Z39.50 (10003)}.

This OID is abbreviated as: ANSI-standard-Z39.50.

Several object classes are assigned at the level immediately subordinate to ANSI-standard-Z39.50, including:

- 3 = attribute set definitions
- 4 = diagnostic definitions
- 5 = record syntax definitions
- 13 = database schema definitions.
- 14 = tagSet definitions.

GILS requires support of the following objects

• GILS attribute set:	{ANSI-standard-Z39.50 3 3}
• bib1 diagnostic set:	{ANSI-standard-Z39.50 4 1}
• USMARC record syntax:	{ANSI-standard-Z39.50 5 10}
• SUTRS record syntax:	{ANSI-standard-Z39.50 5 101}
• GRS-1 record syntax:	{ANSI-standard-Z39.50 5 105}
• GILS schema:	{ANSI-standard-Z39.50 13 2}
• tagSet-M	{ANSI-standard-Z39.50 14 1}
• tagSet-G	{ANSI-standard-Z39.50 14 2}

5.3. Communication Services

Initial implementations of GILS servers will be accessible via the Internet. Therefore, Z39.50 will be using the transport service of the Transmission Control Protocol (TCP). The specification for use of TCP is found in OIW/SIGLA Document #1, "Using Z39.50-1992 Directly over TCP" (Open Systems Environment Implementors Workshop/Special Interest Group on Library Applications (OIW/SIGLA), 1993). The GILS Profile has not defined the use of other communication services.

5.4. Z39.50 Facilities and Services

GILS Z39.50 origins (clients) and targets (servers) must support the following Facilities and Services for information retrieval for operation in the Internet environment:

FACILITY	SERVICE
Init Facility -- allows an origin (client) to propose values for initialization parameters.	Init Service
Search Facility -- enables an origin system (client) to query a database at a target system (server), and to receive information about the results of query.	Search Service
Retrieval Facility -- enables the origin (client) to retrieve records according to position within a result set maintained by the target (server).	Present Service
Termination Facility -- allows the origin (client) or target (server) to initiate abrupt termination or graceful termination of a connection. Mapped to TCP ABORT or TCP CLOSE (see Lynch, 1994 and Open Systems Environment Implementors Workshop/Special Interest Group on Library Applications (OIW/SIGLA), 1993).	

No additional services are required for conformance to the GILS Profile. Other Z39.50 services, however, may be provided optionally by targets (servers) and used by origins (clients).

Standard Z39.50 Init Service negotiation procedures control the use of all services.

5.4.1. Search

This section describes the components and procedures used by Z39.50 to communicate search queries. The GILS application will support Z39.50 Type 1 queries which are general purpose Boolean query structures.

5.4.1.1. Attribute Set

The profile specifies a GILS Attribute Set that is a registered object. The GILS Attribute Set is a superset of the Bib-1 Attribute Set and consists of all Bib-1 Attributes and additional Use Attributes that are defined for GILS elements (see the Application Profile for the Government Information Locator Service (GILS), Annex A: GILS Attribute Set). These newly defined GILS Use Attributes are well-known and correspond semantically to GILS Core Elements.

GILS servers must support a limited number of GILS Attributes. The required GILS Attributes follow. (Note: The GILS Use Attribute is listed followed by the GILS Use Attribute Number and the corresponding GILS Core Element.)

BEST COPY AVAILABLE

- Use Attributes: Local Number (12; Local Control Number); Author-name corporate (1005; Originator); Date/Time Last Modified (1012; Date of Last Modification); Record Source (1019; Record Source); Distributor Name (2001; Distributor Name); Index Terms -- Controlled (2002; Index Terms -- Controlled); Local Subject Index (29; Local Subject Term); Any (1016)

- Structure: Word (2), URx (104), Date (5), Word List (6)

- Relation: Greater than (5), Equal (3).

GILS servers should never return any of the following four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes a combination of these GILS Attributes (see the Application Profile for the Government Information Locator Service (GILS), Annex A: GILS Attribute Set, Table 1 for the recognized and supported combinations of the GILS Attributes).

5.4.1.2. Well-known Search

To facilitate browsing of Locator Records, there will be a well-known search sent by the client to the GILS server. The well-known search consists of the GILS Attribute Set Use Attribute: Local Number; Structure Attribute: URX; and a term of zero length. GILS servers that support browsing of records will create a result set of one or more GILS Locator Records that provide the necessary information to allow clients to offer menu-like displays of GILS Locator Records or other information and information resources.

The "Browse" in the GILS context involves only the Search and Present Services of Z39.50. "Browse" is used informally in the GILS Profile, and it is not related nor should it be confused with the Browse Facility or Scan Service of Z39.50.

5.4.2. Retrieval

This section describes the components and procedures used by Z39.50 to return records in response to a query.

5.4.2.1. Schema

Schemas provide a way to identify the elements that are available from a database record. Each element is defined in a tagSet and is identified by a tagType and a tag value. In addition to describing and/or defining tagSets used in an application, the a schema also includes an abstract record structure (ARS). The ARS describes an abstract structure for a database record, in terms of a set of schema elements, as well as describing the hierarchy of a record.

The GILS Profile specifies a GILS Schema (see the Application Profile for the Government Information Locator Service (GILS), Annex D: GILS Schema). The GILS Schema is a registered object. The schema describes and/or defines tagSets used and an abstract record structure for a Locator Record. A schema in Z39.50 can be modified and may evolve over time, and it is reasonable to expect the GILS Schema will evolve.

The GILS Schema uses tagSet-M and tagSet-G elements and defines in the GILS tagSet additional elements as necessary. The GILS Profile specifies tagTypes to identify tagSet-M elements (tagType = 1), tagSet-G elements (tagType =2), and the elements defined by the GILS tagSet (tagType = 4). Another tagType (tagType=3) is used to identify arbitrary string tags for locally defined elements.

The GILS tagSet element numbering begins with number 1. Elements can be nested and the tagging notation (i.e., the tag path) will reflect the nesting. The form of the notation is (x,y)/(z,w) where x and z are numbers identifying the tagType for the tag and y and w are tag values. For example, for the notation for specifying the element DISTRIBUTOR (4,90) under AVAILABILITY (4,70) would be (4,70)/(4,90).

All well-known GILS Schema elements have assigned numeric tags. String-tags (i.e., text) may be used in the GILS Schema to label those elements that are not well-known (i.e., locally defined).

5.4.2.2. Element Set Names

GILS servers will support four Element Set Names. GILS servers will interpret the use of the Element Set Names required by the GILS Profile to identify the following elements from the GILS Schema:

- The primitive element set name "B" contains at least: title, controlIdentifier, originator, and local control number
- The primitive element set name "G" contains: all B Element Set elements and crossReference
- The primitive element set name "W" contains: all B Element Set elements and bodyOfDisplay
- The primitive element set name "F" contains: all elements available in the record.

The element "bodyOfDisplay" in tagSet-G (2,9) may be used by the target to combine into this single element (i.e., bodyOfDisplay) one or more of the elements from the abstract record structure into a display format.

The server should include in a retrieved record all of the elements specified by the element set name for which there is data available in the database record and which can be encoded in the requested record syntax (e.g., some types of locally defined binary data may not be encodable in a USMARC or SUTRS record).

5.4.2.3. Record Syntaxes

Record syntaxes provide for the transfer of database records between a target (server) and an origin (client) in acceptable form for processing or display.

GILS servers are required to support the following three Z39.50 record syntaxes:

- Generic Record Syntax (GRS-1)
- USMARC
- Simple Unstructured Text Record Syntax (SUTRS).

The Generic Record Syntax is a general-purpose format for packaging records of varying complexity with potentially arbitrary data in individual fields. For mainly text records like GILS Locator Records, GRS-1 is simple and efficient.

USMARC is an implementation of ANSI/NISO Z39.2 and is maintained by the Library of Congress. It is a communications format used by many bibliographic systems. These systems are likely to be important users of GILS. The research team defined a mapping of the GILS Core Elements into the USMARC Format for Bibliographic Data (see the Application Profile for the Government Information Locator Service (GILS), Annex B: GILS Core Elements to USMARC Mapping). However, since the data transformation is not fully reversible and requires interpretation, the record source is responsible for encoding the USMARC record(s).

The data in GILS Locator Records do not always map clearly into USMARC records, particularly when agencies add their own locally defined fields to the GILS Locator Record. This means that construction of USMARC records is subject to local interpretation. Therefore, GILS Locator Records in USMARC format obtained from other than the original record source should be considered non-definitive. The original source of the GILS Locator Record can be identified by examining the ORIGINAL CONTROL IDENTIFIER field in the record.

Unstructured Text (SUTRS) provides a bare-minimum operating capability. SUTRS records consist of a single text field formatted by the target system (server). GILS targets (servers) will use the Preferred Presentation Format (see Section 5.5) to format Locator Records for Unstructured Text transmission.

For interchange, GRS records are to be treated as the complete and canonical representation. SUTRS and USMARC should be viewed as derivative records from the canonical representation and as such are not as complete or precise.

5.5. Preferred Display Format for Use with SUTRS

The GILS Profile recommends a preferred display format for SUTRS records (see the Application Profile for the Government Information Locator Service (GILS), Annex C: Preferred Display Format for GILS Records). For the SUTRS records, formatting instructions for a preferred display format is a concern of the server.

When the target transfers a GILS record using the SUTRS record syntax, it will encode the GILS record formatted according to the preferred display format, so that the client may present the record directly, without processing. For SUTRS, however, the client should not expect to be able to parse the record to obtain any individual GILS elements.

When the client presents a GILS record formatted by the server using the USMARC or GRS record syntax, it is recommended that the client consider the SUTRS suggested display layout in formatting the received record for presentation to the human end user.

5.6. Diagnostic Messages

The GILS application will use Diagnostic Set Bib-1.

6. Data Elements in GILS Locator Records

The GILS document provides the list of data elements for Locator Records. The document refers to these as the GILS Core Elements (see the Application Profile for the Government Information Locator Service (GILS), Annex E: GILS Core Elements, which contains a list of the elements and their definitions).

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The research team has examined the Core Elements and has had input into revisions of these Elements, particularly Elements related to the functional requirements for searching, browsing, and navigating the GILS.

7. Conclusion

This broad outline of the GILS application and the use of Z39.50 in this application is based on the development work of the research project team. During the research project, the team solicited comments from a variety of stakeholders and other interested parties (e.g., the USMARC community, Federal agencies, Z39.50 implementors/vendors, records management and archival community, etc.). Feedback from these groups and other individuals have informed the development of the GILS Profile.

Now that the draft GILS Profile has been completed, the project team will ensure its wide distribution. We anticipate that a number of organizations, companies, vendors, and individuals will develop implementations based on the GILS Profile. A further step in the GILS implementation is a mechanism by which these prototype implementations of the GILS Profile can undergo interoperability testing. Such testing can provide additional feedback on the utility of the GILS Profile, and if necessary, changes and/or expansions to the GILS Profile can be made.

One major goal of this research project has been to ensure that the GILS Profile is implementable and usable, and that implementations based on the Profile can interoperate and interwork. Achieving this goal will serve the larger goals of the Government Information Locator Service by providing a standards-based, decentralized, network-accessible service through which the public will be able to identify and locate Federal information resources. In addition, the GILS Profile provides the means by which various implementors using a variety of computer platforms (clients and servers) can develop products usable by Federal agencies and the public.

NOTES

1. The current draft of Government Information Locator Service (GILS) (Christian, 1994), dated May 2, 1994, is available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at <130.11.48.107> as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

2. The Application Profile for the Government Information Locator Service (GILS) is available via anonymous FTP from <ericir.syr.edu> as /USGS/GILS_PROFILE.ps (Postscript format) or /USGS/GILS_PROFILE.txt (ASCII format).

3. The GILS Profile only addresses the needs of the GILS Core and uses the GILS Core Elements for description used in the GILS Core Locators Records. Throughout this paper, the reader should assume that "GILS" refers to "GILS Core." For further information about the GILS Core, see Christian (1994).

4. Z39.50, Version 2, was approved in 1992. Since that time, the Z39.50 Implementors Group (ZIG), which is a voluntary user group comprising implementors of Z39.50, has continued work to enhance the standard based on needs of information providers. A draft Version 3 is expected to be balloted through the National Information Standards Organization in 1994. The new version of the standard is referred to as Z39.50 -- 1994 and will describe both Version 2 and Version 3 of the standard. Drafts of Version 3 can be retrieved from the Library of Congress's gopher. Connect to MARVEL.LOC.GOV and select #7. Services to Libraries and Publishers, and then select #8. Z39.50.

5. For information on the process by which the objectives, services requirements, and core requirements of GILS were developed, contact Eliot Christian, United State Geological Survey, 802 National Center, Reston, VA 22092; telephone: (703) 648-7245; electronic mail: <echristi@usgs.gov>.

6. USMARC is the implementation in the United States of ANSI Z39.2, the standard for bibliographic information interchange. See American National Standards Institute (1985) and Library of Congress (1993). The Internet Engineering Task Force (IETF) develops standards for the environment of the Internet. For a description of this standards development process see Crocker (1993).

REFERENCES

American National Standards Institute. (1985). American national standard Z39.2-1985: Bibliographic information interchange. New York: American National Standards Institute.

Christian, Eliot. (1994, May 2). Government information locator service (GILS): Report to the information infrastructure task force. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at <130.11.48.107> as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

Crocker, David. (1993, September). Making standards the IETF way. StandardView, 1(1), 48-54.

BEST COPY AVAILABLE

International Organization for Standardization/International Electrotechnical Commission. (1992). ISO/IEC TR10000-1 Information technology — Framework and taxonomy of international standardized profiles — Part 1: Framework. Geneva: ISO/IEC Copyright Office.

Library of Congress. (1993). USMARC format for bibliographic data. Washington, DC: Library of Congress, Cataloging Distribution Service.

Lynch, Clifford A. (1994, April 30). Using the Z39.50 Information Retrieval Protocol in the Internet Environment [Draft RFC for Z39.50 over TCP/IP].

McClure, Charles R., Ryan, Joe & Moen, William E. Moen. (1992). Identifying and describing federal information inventory/locator systems: Design for networked-based locators. 2 Vols. Bethesda, MD: National Audio Visual Center [Available from ERIC, document no. ED349031].

National Information Standards Organization. (1992). ANSI/NISO Z39.50-1992. Information retrieval application service definition and protocol specification for open systems interconnection. Gaithersburg, MD: NISO Press.

Open Systems Environment Implementors Workshop/Special Interest Group on Library Applications (OIW/SIGLA). (1993). OIW/SIGLA Document #1: Using Z39.50-1992 Directly over TCP.

Uniform Resource Locators (URL): A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network. (1993, October). [Internet Draft]. The latest URL draft is: <url:ftp://info.cern.ch/pub/www/doc/url7a.txt>

Uniform Resource Names. (1993, October). [Internet Draft]. The latest URN draft is: <url:ftp://ds.internic.net/internet-drafts/draft-ietf-uri-resource-names-01.txt>

APPENDIX A
Project Team Members

The research project team consists of experts in Z39.50 and also representatives of Federal agencies. The following lists these members:

Z39.50 Experts

Kevin Gamiel	Clearinghouse for Network Information Discovery and Retrieval
Ralph LeVan	OCLC
Denis Lynch	ESL, Inc.
Margaret St. Pierre	WAIS, Inc.
Madeleine Stovel	Research Libraries Group, Inc.

Representatives from Federal Agencies

Eliot Christian	United States Geological Survey
Tim Gauslin	United States Geological Survey
Sue Ruddle	Defense Technical Information Center
Yesha Yelena	National Institute of Standards and Technology

Representative of the Z39.50 Maintenance Agency

Ray Denenberg	Z39.50 Maintenance Agency, Library of Congress

APPENDIX B
Definitions

For purposes of this Profile, the following definitions apply.

Association: A communication session between a database user and a database provider.

Client: An initiating application. This application includes the Z39.50 origin.

Electronic Information Resource: Information resources that are maintained in electronic, digital format and may be accessed, searched, or retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM).

GILS Core: A subset of all GILS Locator Records which describe information resources maintained by the U.S. Federal government, all of which comply with the defined GILS Core Elements and are mutually accessible through interconnected electronic network facilities without charge to the direct user.

Government Information: Information created, collected, processed, disseminated, or disposed of by or for the Federal government.

Government Information Locator Service (GILS) : A decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

Information: Any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.

Information Resource: Includes both government information and information technology.

Interoperability: A condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

Locator: An information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

Locator Record: A collection of related data elements describing an information resource, the information available in the resource, and how to obtain the information. Locator Records will be offered by servers to identify information resources, describe the information available in those resources, and provide assistance in how to obtain the information.

Mandatory: An element in a GILS Core Locator Record that must have a value provided by the record source. The GILS Profile does not specify which elements must be present from the perspective of GILS servers.

Origin: The part of a client application that initiates a Z39.50 association and is the source of requests during the association.

Profile: The statement of a function(s) and the environment within which it is used, in terms of a set of one or more standards, and where applicable, identification of chosen classes, subsets, options, and parameters of those standards. A set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

Registered Object: An object that is identified by a name-to-thing relationship in which the name is recorded by a registration authority to ensure that the names can be used unambiguously.

Server: An application that responds to an initiating application (i.e., a client). The application that includes the Z39.50 target.

Target: The part of an server application that accepts a Z39.50 association.

Uniform Resource Identifier (URI): A set of related standards for encoding resource location and identification information for electronic and other objects. Examples include Uniform Resource Locators (URLs) and Uniform Resource Names (URNs).

USMARC: An implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data. This documentation is published by the Library of Congress.

Attachment D

USING THE Z39.50 INFORMATION RETRIEVAL PROTOCOL IN THE INTERNET ENVIRONMENT

IIIR Working Group
INTERNET-DRAFT
<draft-ietf-iiir-z3950-01.txt

Clifford Lynch
University of California
Office of the President

Using the Z39.50 Information Retrieval Protocol
in the Internet Environment

Status of This Document

This memo describes an approach to the implementation of the ANSI/NISO Z39.50-1992 Standard for Information Retrieval in the TCP/IP environment which is currently in wide use by the Z39.50 implementor community.

This memo provides information for the Internet community. This memo does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

> This document is an Internet Draft. Internet Drafts are working documents of the Internet Engineering Task Force (IETF), its Areas, and its Working Groups. Note that other groups may also distribute working documents as Internet Drafts.
>
> Internet Drafts are draft documents valid for a maximum of six months. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a "working draft" or "work in progress."
>
> Please check the I-D abstract listing contained in each Internet Draft directory to learn the current status of this or any other Internet Draft.
>
> This Internet Draft expires February 1, 1995.

Summary

This memo describes an approach to the implementation of the ANSI/NISO Z39.50-1992 Standard for Information Retrieval in the TCP/IP environment which is currently in wide use by the Z39.50 implementor community.

Introduction

Z39.50 is a US national standard defining a protocol for computer-to-computer information retrieval that was first adopted in 1988 [1] and extensively revised in 1992 [2]. It was developed by the National Information Standards Organization (NISO), an ANSI-accredited standards development body that serves the publishing, library, and information services communities. The closely related international standard, ISO 10162 (service definition) [3] and 10163 (protocol) [4], colloquially known as Search and Retrieve or SR, reached full International Standard (IS) status in 1991. Work is ongoing within ISO Technical Committee 46 Working Group 4 Subgroup 4 to progress various

extensions to SR through the international standards process. The international standard is essentially a compatible subset of the current US Z39.50-1992 standard. Z39.50 is an applications layer protocol within the OSI reference model, which assumes the presence of lower-level OSI services (in particular, the presentation layer [5]) and of the OSI Association Control Service Element (ACSE) [6] within the application layer.

Many institutions implementing this protocol chose, for various reasons, to layer the protocol directly over TCP/IP rather than to implement it in an OSI environment or to use the existing techniques that provide full OSI services at and above the OSI Transport layer on top of TCP connections (as defined in RFC 1006 [7] and implemented, for example, in the ISO Development Environment software). These reasons included concerns about the size and complexity of OSI implementations, the lack of availability of mature OSI software for the full range of computing environments in use at these institutions, and the perception of relative instability of the architectural structures within the OSI applications layer (as opposed to specific application layer protocols such as Z39.50 itself). Most importantly, some of these institutions were concerned that the complexity introduced by the OSI upper layers would outweigh the relatively meager return in functionality that they were likely to gain. Thus, for better or worse, the decision was taken to implement the Z39.50 protocol directly on top of TCP (with the understanding that this decision might be revisited at some point in the future).

During 1991-1993, a group of implementing institutions agreed to participate in the Z39.50 Interoperability Testbed project (sometimes referred to by the acronym "ZIT") under the auspices of the Coalition for Networked Information (CNI). Their primary objective was to encourage the development of many interoperable Z39.50 implementations running over TCP/IP on the Internet. By mid-1993, a number of independent Z39.50 implementations were operational and able to interoperate across the Internet.

The Library of Congress, in its role as the Z39.50 Maintenance Agency for NISO, maintains a registry of the implementors [8], which includes members of the Z39.50 interoperability testbed.

This document describes implementation decisions by current implementors of Z39.50 in the Internet environment. These have been proven within the ZIT project and are being used by most of the members of the Z39.50 Implementors' Group (ZIG), an informal group that meets quarterly to discuss implementation and interoperability issues and to develop extensions to the Z39.50 protocol targeted for inclusion in future versions of the standard. Intended as a guide for other implementors who seek to develop interoperable Z39.50 implementations running over TCP/IP, this document focuses on issues related to TCP/IP, and it does not address other potential interoperability problems or agreements that have been reached among the implementors to address these problems. It does include a few notes about extensions to the existing Version 2 protocol that are being used in the implementor community which have interoperability implications. Potential implementors of Z39.50 should subscribe to the Z3950IW LISTSERV [9] to obtain information specific to the Z39.50 protocol and extensions under development as well as details of current implementations.

Except where otherwise noted, the version of Z39.50 discussed here is

ANSI/NISO Z39.50-1992, sometimes called Z39.50 Version 2 (the obsolete original version is referred to as Z39.50-1988 or Z39.50 Version 1). The approach defined should also be applicable, perhaps with some minor changes, to future versions of the Z39.50 protocol, and specifically to Version 3 which is currently under development. This document will probably be updated to address new versions of the base Z39.50 protocol as they become stable.

Encoding

The Z39.50 standard specifies its application protocol data units (APDUs) in Abstract Syntax Notation One (ASN.1) [10]. These APDUs include EXTERNAL references to other ASN.1 and non-ASN.1 objects such as those defining record transfer syntaxes to be used in a given application association.

The standard Basic Encoding Rules (BER) [11] are applied to the ASN.1 structures defined by the Z39.50 protocol to produce a byte stream that can be transmitted across a TCP/IP connection. The only restriction on the use of BER to produce this byte stream is that direct, rather than indirect, references must be used for EXTERNAL objects. This is necessary because there is no presentation context in the TCP/IP environment to support indirect reference. A Z39.50 implementation developed according to this specification and running over TCP/IP should produce a valid byte stream according to the Z39.50 standard, in the sense that the same byte stream could be passed to an OSI implementation. However, not all byte streams that can be produced by applying BER to the APDUs specified in the Z39.50 standard in an OSI environment will be legitimate under this specification for the TCP/IP environment; this specification defines a subset of the possible byte streams valid in a pure OSI environment which excludes those using indirect reference for EXTERNAL objects.

All other BER features should be tolerated by Z39.50 implementations running over TCP/IP, including the ability to accept indefinite length encodings, although it is preferable that implementations do not generate such encodings since they have caused problems for some ASN.1/BER parsers. It should also be noted that at least to the best of the author's knowledge, there are no implementations at present that use ASN.1/BER representations of floating point numbers; instead, integers with scaling factors have been used for these purposes. It should also be noted that Z39.50 version 2 does not really address character set encoding issues; these questions, and their interactions with ASN.1/BER support for multiple character sets, are under active discussion as part of the effort to develop Z39.50 version 3.

Connection

In the Internet environment, TCP Port 210 has been assigned to Z39.50 by the Internet Assigned Number Authority [12]. To initiate a Z39.50 connection to a server in the TCP/IP environment, a client simply opens a TCP connection to port 210 on the server and then, as soon as the TCP connection is established, transmits a Z39.50 INIT APDU using the BER encoding of that INIT APDU as described above.

Implementors should be aware that there is a substantial installed base of implementations of the Wide Area Information Server (WAIS) system. The

original versions of this software employed Z39.50 Version 1 with some extensions. Z39.50 Version 1 did not use BER encoding and Z39.50 Version 1 INIT APDUs look very different from the INIT APDUs of Z39.50 Version 2. Implementations of Z39.50 should at least be prepared to reject gracefully WAIS-type INIT APDUs. Some implementations recognize such INIT APDUs and revert to the Z39.50 Version 1 variant used in WAIS upon encountering them, thus providing backwards compatibility with the existing base of WAIS clients and; the usual means of checking for a WAIS, as opposed to Z39.50 Version 2, client is to see if the first byte sent on the connection is an ASCII zero, which indicates a WAIS client. (In version 1 of WAIS, bytes 0-9 of the first PDU contain an ASCII packet length; the lower case ASCII string "wais" appears starting at byte 12.) Work is currently underway to specify a WAIS profile for use with Z39.50 version 2 [13]; it is expected that this will be issued as a Z39.50 Applications Profile through the NIST OIW Library Automation Special Interest Group. This profile is expected to be compatible with the layering defined in this RFC.

Service Mappings

The Z39.50 standard maps Z39.50 services onto a variety of association control and presentation layer services. Connection establishment has already been discussed. The other two association control services that are relevant to Z39.50 are ABORT and RELEASE. The mapping of the RELEASE service to a standard TCP CLOSE is straightforward. The Z39.50 protocol itself does not, in the current version, include a Z39.50 CLOSE APDU. When the client has completed its interaction with the server, it calls the IR-RELEASE service, which is directly mapped to association control's orderly association release. In the TCP/IP environment, the client should simply initiate a TCP CLOSE. The mapping for association abort is more complex, partially because some TCP/IP implementations cannot distinguish a TCP reset from the other side of the connection from other events. To accomplish an abort (that is, a mapping of the IR-ABORT service in the Z39.50 protocol) in the TCP/IP environment, client or server need only terminate the TCP connection either via TCP ABORT or TCP CLOSE. Real-world implementations need to be prepared to deal with both TCP ABORT and CLOSE anyway, so this approach presents no additional problems, other than the somewhat ambiguous nature of the type of association termination.

It is expected that Z39.50 Version 3 will include a termination service which will involve an exchange of Z39.50 CLOSE APDUs, followed by an association RELEASE (which would presumably, in the Internet environment, be mapped to a TCP CLOSE). This new termination service is expected to support both graceful and abrupt termination. Of course, robust implementations will still need to be prepared to encounter TCP CLOSE or ABORT.

Service mappings for the transmission of data by client and server (to the presentation layer P-DATA service) are trivial: They are simply mapped to TCP transmit and receive operations. TCP facilities such as expedited data are not used by Z39.50 in a TCP environment.

Contexts

At the point when the TCP connection is established on TCP port 210, client and server should both assume that the application context given in Appendices A and B of the Z39.50-1992 standard are in place. These are the

ASN.1 definitions of the Z39.50 APDUs and the transfer syntax defined by applying the BER to these APDUs.

Implementations can reasonably expect that the diagnostic set BIB-1 is supported, and, if resource control is being used, the resource report format BIB-1 is supported as well.

In the absence of a presentation negotiation mechanism, clients and servers should be cautious about using alternative attribute sets, diagnostic record formats, resource report formats, or other objects defined by optional EXTERNALs within the Z39.50 ASN.1, such as authentication parameters, unless there is known to be prior agreement to support them. Of course, either participant in an association can reference such an object by object ID in an APDU, but there is no guarantee that the other partner in the association will be able to understand it. Robust implementations should be prepared to encounter unknown or unsupported object IDs and generate appropriate diagnostics. Over time, the default, commonly known pool of object IDs may be expanded (for example, to support authentication parameters).

Implementors should refer to the document [14] issued by the Z39.50 maintenance agency in June 1992 for more details on the assumed contexts and object identifiers.

Record syntaxes present a serious practical problem. In the OSI environment, the partners in a Z39.50 association are assumed to agree, either through presentation negotiation as part of association establishment, or later, dynamically, as part of the PRESENT process (through the use of the alter presentation context function at the presentation layer), on which record syntaxes the two entities commonly know. There is a preferred record syntax parameter that can be supplied by the client to guide this negotiation. A number of registered record syntaxes exist; some are based on ASN.1 and others use formats such as the MARC standard for the interchange of machine readable cataloging records which predate ASN.1, but are widely implemented. In the TCP/IP environment, if the server cannot supply the record in the preferred syntax, it has no guarantee that the client will understand any other syntax in which it might transmit the record back to the client, and has no means of negotiating such syntaxes.

Several proposals have been suggested to solve this problem. One, which will likely be part of Z39.50 Version 3, is to replace the preferred record syntax parameter with a list of prioritized preferred syntaxes supplied by the client, plus a flag indicating whether the server is allowed to substitute a record syntax not on the list provided by the client. The currently proposed ASN.1 for this extension is upwards compatible with Z39.50 Version 2, although the details are still under discussion within the Z39.50 Implementor's Group. As the Version 3 ASN.1 becomes stable in this area, Z39.50 servers are encouraged to accept the extended ASN.1 for generalized preferred record syntax. The extensibility rules for Z39.50 negotiation let clients and servers negotiate the use of Z39.50 Version 2 plus the generalized preferred syntax feature from Version 3. Thus, a client could support the generalized preferred record syntax, propose its use to any server, and, if the server rejects the proposal, revert to the Version 2 preferred syntax feature.

A second alternative (not incompatible with the Version 3 extension) would be to adopt a convention for TCP/IP implementations that the server not return a record in a syntax not on the preferred record syntax list provided by the client. Instead, it would return a diagnostic record indicating that a suitable record transfer syntax was not available. This strategy could be viewed as simply implementing a subset of the Version 3 solution, and should be considered by implementors of servers as a possible interim measure.

Other Interoperability Issues

Version 3 will include an "other" data field in each APDU, which can be used to carry implementation-specific extensions to the protocol. A number of implementations are already employing this field, and interoperable implementations might be wise to include code which at least ignores the presence of such fields rather than considering their presence an error (in contravention of the standard).

Security Considerations

This document does not discuss security considerations. However, it should be noted that the Z39.50 protocol includes mechanisms for authentication and security that implementors should review.

Author's Address

Clifford A. Lynch
University of California, Office of the President
300 Lakeside Drive, 8th Floor
Oakland, CA 94612-3550
(510) 987-0522
clifford.lynch@ucop.edu

References

[1] National Information Standards Organization (NISO). American National Standard Z39.50, Information Retrieval Service Definition and Protocol Specifications for Library Applications (New Brunswick, NJ: Transaction Publishers; 1988).

[2] ANSI/NISO Z39.50-1992 (version 2) Information Retrieval Service and Protocol: American National Standard, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection, 1992.

[3] ISO 10162 International Organization for Standardization (ISO). Documentation -- Search and Retrieve Service Definition, 1992.

[4] ISO 10163 International Organization for Standardization (ISO). Documentation -- Search and Retrieve Protocol Definition. 1992

[5] ISO 8822 - Information Processing Systems - Open Systems Interconnection - Connection Oriented Presentation Service Definition, 1988

[6] ISO 8649 - Information Processing Systems - Open Systems Interconnection - Service Definition for the Association Control Service Element, 1987. See also ISO 8650 - Information Processing Systems - Open Systems Interconnection - Protocol Specification for the Association Control Service Element, 1987.

[7] RFC 1006, ISO Transport Layer Services on Top of the TCP, Version 3, Rose, Marshall and Dwight Cass, May 1987

[8] Registry of Z39.50 Implementors, available from the Z39.50 Maintenance Agency (Ray Denenberg, ray@rden.loc.gov)

[9] To subscribe to the Z39.50 Implementor's Workshop list send the message
SUB Z3950IW yourname
to
LISTSERV@NERVM.NERDC.UFL.EDU (or NERVM.BITNET).

Current drafts of the Version 3 Protocol document are available through the Library of Congress GOPHER server, MARVEL.LOC.GOV.

[10] ISO 8824 - Information Processing Systems - Open Systems Interconnection - Specifications for Abstract Syntax Notation One (ASN.1), 1987

[11] ISO 8825 Information Processing Systems - Open Systems Interconnection - Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1) 1987

[12] RFC 1340 Assigned Numbers. Reynolds, J.K.; Postel, J.B. July 1992

[13] WAIS Profile of Z39.50 Version 2, Revision 1.4, April 26, 1994, available from WAIS Inc.

[14] Registration of Z39.50 OSI Object Identifiers (Z39.50-MA-024), available from the Z39.50 Maintenance Agency (Ray Denenberg, ray@rden.loc.gov)

The Government Information Locator Service (GILS)

Report to the Information Infrastructure Task Force
May 2, 1994

The Government Information Locator Service (GILS)

Executive Summary

In coordination with the Information Infrastructure Task Force (IITF), the Office of Management and Budget (OMB) is promoting the establishment of an agency-based Government Information Locator Service (GILS) to help the public locate and access information throughout the Federal Government. This report presents a vision of how GILS will be implemented.

Working primarily with OMB and the Locator Subgroup of the Interagency Working Group on Public Access, Eliot Christian of the U.S. Geological Survey prepared this report under the auspices of the IITF Committee on Information Policy. This vision of GILS has also received extensive review by various Federal agencies and other interested parties, including some non-Federal organizations and by the general public through notices in both the Federal Register and the Commerce Business Daily and at a public meeting held in December, 1993.

As part of the Federal role in the National Information Infrastructure, GILS will identify and describe information resources throughout the Federal government, and provide assistance in obtaining the information. It will be decentralized and will supplement other agency and commercial information dissemination mechanisms.

The public will use GILS directly or through intermediaries, such as the Government Printing Office, the National Technical Information Service, the Federal depository libraries, other public libraries, and private sector information services. Direct users will have access to a GILS Core accessible on the Internet without charge. Intermediate access may include kiosks, "800 numbers," electronic mail, bulletin boards, FAX, and off-line media such as floppy disks, CD-ROM, and printed works.

GILS will use standard network technology and the American National Standards Institute Z39.50 standard for information search and retrieval so that information can be retrieved in a variety of ways. Direct users will eventually have access to many other Federal and non-Federal information resources, linkages to data systems, and electronic delivery of information products.

Development of this report proceeded in tandem with a GILS Profile development project that produced an Implementors Agreement in the voluntary standards process. The National Institute of Standards and Technology is now establishing a Federal Information Processing Standard referencing the GILS Profile Implementors Agreement and making mandatory its application for Federal agencies establishing locators for government information.

Existing law and policy, as articulated in OMB Circular A-130, the Records Disposal Act, and the Freedom of Information Act, require agencies to create and maintain an inventory of their information systems and information dissemination products. Although compliance with these requirements varies greatly, the incremental cost of making those inventories accessible through GILS is expected to be minimal. Accordingly, participation in establishing and maintaining GILS may be accomplished as a collective effort executed within existing funds and authorities. OMB will publish in 1994 a Bulletin following on Circular A-130 that will specify agency responsibilities in GILS and set implementation schedules. A process for ongoing evaluation will also be established to evaluate the degree to which GILS meets the information needs of the public.

The Government Information Locator Service (GILS)

Introduction

Government information is fundamental to modern societies. Although individual Federal agencies may recognize their responsibility to maintain readily accessible inventories of their records and other information resources, there needs to be a collective vision across the Federal government for information dissemination to the public. The vision of a Government Information Locator Service (GILS) presented here responds to that need and places this Federal vision in the context of broader issues such as promotion of diverse information services.

GILS is emerging at a revolutionary period in the history of information processing where technological breakthroughs have radically expanded the range of feasible strategies. In particular, the realization of peer computer networks allows for a decentralized approach where many different information sources are separately maintained yet are comprehensible as a coherent whole from the unique perspective of a specific user. GILS depends on this network approach to preserve the decentralized character of Federal information dissemination and the wide diversity of sources, both public and private, that serve the public need for information access.

In contrast to a centralized design, a decentralized approach assumes that many different implementations will be separately developed yet will be fully interoperable when implemented. Achieving interoperability is only possible if a stable base of reference is documented and made widely known. In GILS, that reference base is an agreement among active implementors together with Federal representatives. Where fundamental design choices have been made in developing the implementors agreement, those choices have emphasized the use of stable but extensible standards.

The success of GILS does not depend on massive Federal investment or sweeping new directives. Rather, it adopts voluntary information standards in order to build on the efforts of the responsible, talented, and creative people throughout Government and in society already working on information access issues. GILS will use this solid base of widely accepted standards to help agencies and information services focus their initiatives and thereby make the vast range of Government information more accessible to the public.

Policy Context

The Administration's strategic technology policy document entitled "Technology for America's Economic Growth, A New Direction to Build Economic Strength" states:

> Every year, the Federal Government spends billions of dollars collecting and processing information (e.g., economic data, environmental data, and technical information). Unfortunately, while much of this information is very valuable, many potential users either do not know that it exists or do not know how to access it. We are committed to using new computer and networking technology to make this information more accessible to the taxpayers who paid for it. In addition, it will require consistent Federal information policies designed to ensure that Federal information is made available at a fair price to as many users as possible while encouraging growth of the information industry.[1]

On June 25, 1993, the Office of Management and Budget (OMB) revised Circular A-130, "Management of Federal Information Resources," to strengthen policies for managing government information (58 F.R. 36068, July 2, 1993). Circular A-130 encourages agencies to use new technologies to make government information available to the public in a timely and equitable manner via a diverse array of sources, both public and private. It states that availability of government information in diverse media, including electronic formats, permits the public greater flexibility in using the information, and that modern information technology presents opportunities to improve the management of government programs to provide better service to the public. It also notes that the development of public electronic information networks, such as the Internet, provides an additional way for agencies to increase the diversity of information sources available to the public, and that emerging standards such as ANSI (American National Standards Institute) Z39.50[2] will be used increasingly to facilitate dissemination of government information in a networked environment.

OMB Circular A-130 states that agencies shall:

- Disseminate information products on equitable and timely terms;
- Avoid establishing, or permitting others to establish on their behalf, exclusive, restricted, or other distribution arrangements that interfere with the availability of information dissemination products on a timely and equitable basis;
- Use voluntary standards and Federal Information Processing Standards where appropriate or required;
- Use electronic media and formats, including public networks, as appropriate and within budgetary constraints, in order to make government information more easily accessible and useful to the public;

---

[1] Clinton, William J. & Gore, Albert, Jr., (1993, February 22). Technology for America's Strength, A New Direction to Build Economic Strength. Washington, DC: Government Printing Office.

[2] National Information Standards Organization. (1992). ANSI/NISO Z39.50-1992, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. Gaithersburg, MD: National Information Standards Organization Press.

- Take advantage of all dissemination channels, Federal and nonfederal, including State and local governments, libraries and private sector entities;
- Provide information describing how the public may gain access to agency information resources;
- Help the public locate government information maintained by or for the agency;
- Establish and maintain inventories of all agency information dissemination products;
- Develop such other aids to locating agency information dissemination products including catalogs and directories...

Because the active management of information by agencies is essential to the operation of government and to democratic principles, laws and policies assert a fundamental requirement that Federal agencies maintain readily accessible inventories of their records and other information holdings. The responsibilities of Federal agencies with regard to the management of electronic records are also growing in importance as their reliance on electronic information systems increases. To help the public locate and gain access to public information within agency inventories, the Administration has committed to promote the establishment of an agency-based Government Information Locator Service (GILS).

Working primarily with OMB and the Locator Subgroup of the Interagency Working Group on Public Access (the "Solomon's Group"), Eliot Christian of the U.S. Geological Survey (USGS) prepared this report to the Information Infrastructure Task Force describing how GILS may be implemented. Development of this report proceeded in tandem with a GILS Profile development project that produced an Implementors Agreement in the voluntary standards process. The GILS Profile project was a Cooperative Agreement between the USGS and Syracuse University, funded by the Interagency Working Group on Data Management for Global Change, with active involvement from several ANSI Z39.50 implementors representing non-government sectors.[3] The National Institute of Standards and Technology (NIST) is now establishing a Federal Information Processing Standard (FIPS) referencing the GILS Profile Implementors Agreement and making mandatory its application for Federal agencies establishing locators for government information.

Existing law and policy, as articulated in OMB Circular A-130, the Records Disposal Act (Title 44 of the United States Code), and the Freedom of Information Act (FOIA), already require agencies to create and maintain an inventory of their information systems and information dissemination products. Although compliance with these requirements varies greatly, the incremental cost of making those inventories accessible through GILS is expected to be minimal. Accordingly, participation in meeting the minimum mandatory requirements for establishing and maintaining GILS may be accomplished as a collective effort within existing funds and authorities.

OMB will publish in 1994 a Bulletin following on Circular A-130 that will specify agency responsibilities in GILS and set implementation schedules. A process for ongoing evaluation will also be established to evaluate the degree to which GILS meets user information needs, including

---

[3] McClure, Charles R., & Moen, William E. (1994). Expanding Research and Development on the ANSI/NISO Z39.50 Search and Retrieval Standard. Syracuse, NY: School of Information Studies, Syracuse University.

factors such as accessibility, ease of use, suitability of descriptive language, accuracy, consistency, timeliness, and completeness of coverage.

The User Perspective

GILS must be many things to many people. It must be comprehensive, yet user friendly. It must answer specific questions, yet enable scanning a wide range of government information. It must be able to answer questions from the most inexperienced users, yet permit in-depth research as well. It must be of direct service to the public, yet not undermine the diversity of existing information sources. Private-sector information providers must be able to participate in GILS and also make their resources known and accessible.

GILS depends critically on other aspects of the emerging NII. GILS must be implemented with full recognition of individual privacy and intellectual property rights. Agencies will need to ensure that members of the public whom the agency has a responsibility to inform have a reasonable ability to access GILS and the underlying information resources and information dissemination products. Agencies participating in GILS must take care to minimize barriers to use, including equipment and software requirements, cost, and technical complexity.

The public will use GILS either directly or through intermediaries. The distinction is that direct users roam at will, but users of intermediate services take a guided tour. The following are some examples of GILS direct users and intermediaries:

- A direct user researching national health care may explore relevant issues from a variety of perspectives by accessing a wide range of GILS and non-GILS information sources.
- An educator interested in keeping up with electronic educational materials may access a few GILS sources once a month as a direct user over a dial-up connection to the Internet
- An information service may query GILS hourly as a direct user and also act as an intermediary by constructing a value-added directory derived from GILS for sale to users who need specific products such as government economic statistics.
- A Federal agency may act as an intermediary in adding GILS access into its existing information service to provide public information referrals to sources in other agencies.

A major advantage of the networked and decentralized design of GILS is that it allows direct users to explore many different aspects of government information. Since direct users are less limited in their searching, they have more flexibility to explore the full complement of available information. For direct users, there is minimal structure across the GILS locator records and the records are interleaved with a vast diversity of other kinds of information. On the Internet, direct users have tools for interacting with people, news, and libraries in addition to GILS (Figure 1).

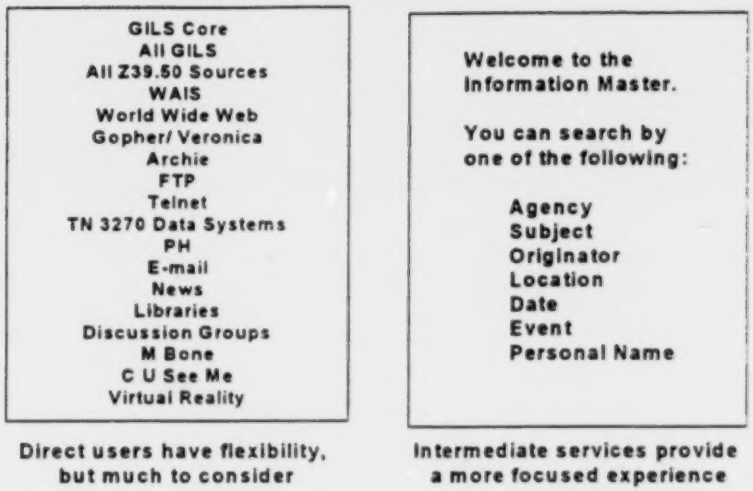

Figure 1. The public will use GILS either directly or through intermediaries.

In contrast, intermediate services are typically oriented toward a particular user community and present a more focused experience for users searching for information. Intermediate services need not require users to have sophisticated research skills or electronic network access. Government and non-government intermediaries can present GILS information in the full range of communications media and with a variety of interpretative services as appropriate for various communities. Such services can be offered via electronic mail, bulletin boards, FAX, and other media such as CD-ROM (Compact Disk-Read Only Memory), printed publications, telephone help desks, and information kiosks in public places as envisioned in the Administration's Service to the Citizen initiative.[4]

Clearly, most of the public need for access to government information will be well served through the diverse array of public and private-sector service providers. Casual users and those lacking network access will be served typically through products and services offered by agency or non-government intermediaries such as Federal depository libraries, other public libraries, and private-sector providers. These intermediaries obtain GILS information either as direct users themselves or from other intermediaries, but the extent of government information that may be provided by any particular intermediate service is not prescribed by GILS.

Having unfettered access means that the direct user takes on much more responsibility to construct a context in which the collected information is actually coherent. Accordingly, GILS has certain expectations of direct users, whether researchers or other intermediaries. Direct users of GILS must have network access, be literate in English to at least the secondary-school level, be capable of using a personal computer, and be aware of any limitations of their own hardware or software environment.

---

[4] Service to the Citizen Interagency Task Force. (1993). Service to the Citizen Conference Report. Washington, DC: Department of Veterans Affairs.

Data and Information

Given the huge amounts and vast range of Federal holdings, one might want to synthesize information by combining data from multiple sources as, for example, to support large scale environmental monitoring. It is important to understand that GILS operates at the level of information about data holdings. GILS addresses how to find files but does not address how the contents of those files may be accessed or used.

Users must be aware that data combined from multiple sources should be used with caution and subjected to appropriate review. Except in very strictly defined domains where common practices are rigidly enforced and data processing is well coordinated, there does not exist sufficiently detailed documentation about the data to ensure its appropriate use for purposes other than for which it was initially gathered. This situation is not peculiar to Federal holdings--whenever data is collected and maintained, it is only possible to provide for a limited set of secondary uses.

In some communities of interest, such as the participants in the National Spatial Data Infrastructure, there is strong consensus on the high secondary use value of certain basic data. This perceived value justifies large investments in data management and the establishment of multi-lateral coordination structures such as the Federal Geographic Data Committee established under OMB Circular A-16. Data management issues surrounding the international Global Change Research Program and the work of the Committee on Earth and Natural Resources are also generating some convergence of opinion on raising the level of data management investments.

While there are complex issues surrounding data comparability, it is clear that complete and readily accessible information about data holdings will be a key requirement. GILS does provide a basis for broad accessibility to the highest level description of information holdings.

The Provider Perspective

A key concept of GILS is that it uses network technology to support many different views across many separate locators.[5] A locator is defined as an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

Although directly accessible via electronic networks such as the Internet, all or part of the GILS contents can also be made available by intermediaries through virtually any media. These alternative mechanisms help assure that the information is available through a diversity of sources, both public and private, and cover the full range of communications media from telephone help though printed publications and up to the most sophisticated electronic network technologies.

GILS organizes a collective set of agency-based locators and associated information services. Being decentralized, responsibilities can be kept close to those who understand and care for the information and who are serving the agency's primary user community. Each agency is responsible

---

[5] The design of GILS follows generally a 1992 report to OMB, NARA, and the General Services Administration (GSA): McClure, Charles R., Ryan, Joe & Moen, William E. (1992). Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-based Locators 2 Vols. Bethesda, MD: National Audio Visual Center. [Available from ERIC, document no. ED349031].

for ensuring that its GILS components are continuously accessible to GILS direct users. Certain agencies, such as NARA, the Government Printing Office (GPO), and the National Technical Information Service (NTIS), also have in their primary mission an additional role in helping the public to access information maintained elsewhere in the Government. These agencies will assist in providing GILS services when requested by other agencies.

Services for finding government information take many forms, and the electronic aspects of GILS should be seen within the larger context of government information services (Figure 2). For example, the public is served through information desks in Federal buildings as well as telephone help desks and reference services such as "1-800-USA-MAPS." Many kinds of finding aids are used in such services--printed catalogs and directories are and will continue to be very common. With GILS, it will be much easier for those services to provide information drawing on the full range of Federal information resources rather than just agency-specific resources.

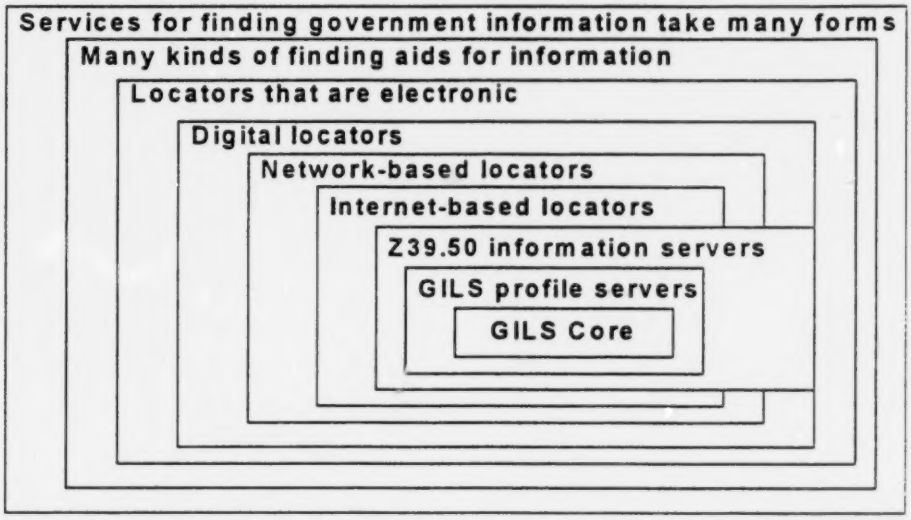

Figure 2. Electronic networks are one aspect of the Government Information Locator Service.

Among the government information finding aids are electronic media, including television announcements about government information available from the Consumer Information Center in Pueblo, Colorado. As interactive television becomes more available to homes, GILS will help to simplify the ways in which those services help the pubic to find Federal information resources. Also within the realm of digital electronic finding aids, there are popular information dissemination technologies such as bulletin boards and CD-ROM's. These personal, print media, and electronic services can be used to publicize GILS contents. These services may also be regarded as information resources, and may be referenced in GILS locator records themselves.

Some digital electronic finding aids use various kinds of networks and so are able to provide access to many different resources, often with a common user interface. In this area, it becomes possible to provide services in GILS where the user can have immediate access not only to information about an information resource, but to the referenced resource itself.

As stated above, GILS takes advantage of network technologies to allow many different information sources to be separately maintained yet be comprehensible as a coherent whole from the unique perspective of a specific user. This is achieved within computer networks that support peer-to-peer relationships and thereby allow for applications to operate using a client-server architecture. All of the server applications that also use the ANSI Z39.50 information search and retrieval protocol can be accessed by GILS direct users.

Because GILS uses interoperable standards for information search and retrieval, information sources referenced in GILS can be placed into virtually any context. Other major Federal government information systems such as the GPO Access System, the NTIS FedWorld system, the National Geospatial Data System, and the Global Change Data and Information System will be accessible to GILS direct users. GILS direct users may have access to a wide range of additional Federal information on the network such as current and historical information on Federal programs and institutions; public notices; law, regulation, policy, and procedural materials; and listings of experts and office locations. Agencies such as NARA, GPO, and NTIS, as well as private-sector information providers, can supplement the GILS Core with access to other Federal and non-Federal information.

Other government (state, local, tribal, foreign, international) and non-government organizations will also be encouraged to institute locators compatible with the standards used in GILS. GILS will accommodate the expressed needs of other government organizations where practical.

Design Principles

GILS is a component of the National Information Infrastructure (NII) that is evolving with guidance from the Information Infrastructure Task Force.[6] GILS will be interoperable with other component NII initiatives such as the National Spatial Data Infrastructure. GILS is also expected to adapt to and encourage technical innovation, especially in ways that enhance public access to government information.

GILS will conform to national and international standards for information and data processing. Participants in GILS will use voluntary standards processes, e.g., ANSI, the Open Systems Environment Implementors Workshop (OIW), and the Internet Engineering Task Force, to promote interoperability of search and retrieval mechanisms, network communications, user authentication, and resource identifiers, among other essential components. Near-term implementations of GILS will use the Internet and its communications protocols, but GILS is based on the international Open Systems Interconnection (OSI) model to be compatible with a wide range of technologies. NIST, working through the OIW, will maintain and publish the application profile specifying GILS compliance.

---

[6] Information Infrastructure Task Force (September 15, 1993). The National Information Infrastructure: Agenda for Action. Washington, DC: NTIA NII Office, Department of Commerce. Available in ASCII text format under the file name niiagend.asc on the NTIA Bulletin Board (202) 482-1199 and the Fedworld bulletin board (703-321-8020). It is available on the Internet under the file name niiagenda.asc by anonymous FTP (File Transfer Protocol) at host ftp.ntia.doc.gov under the directory /pub, and by gopher at gopher.nist.gov in the menu item DOC Documents.

GILS takes advantage of the network technology known as client-server architecture, which allows locator records to be distributed among multiple independent information servers. Client applications may allow the user to question many servers concurrently and have the answers automatically combined. In this way, GILS allows for agencies to maintain GILS locator records within various information resources optimized for their usual customers, while allowing the locator information to be rapidly collated in different ways to serve different needs.

Functional Requirements

Direct users of GILS must be able to use non-proprietary standard mechanisms to discover information sources and retrieve basic textual information content. These functions are within the scope of the information search and retrieval standard known in the United States as ANSI Z39.50 and internationally as ISO (International Organization for Standardization) 10162/10163. GILS locators must be accessible on interconnected electronic network facilities and must support the currently approved ANSI Z39.50 standard for information search and retrieval. Software conforming with ANSI Z39.50 must also conform to the GILS Profile to provide full functionality to GILS direct users. In particular, the GILS Profile provides for navigating among Federal government locators through the specifications given for the GILS Core locator records. Special provisions are made in GILS to support navigation among GILS locators by using browsing as well as textual searching.

The GILS Profile provides a complete specification of GILS as it makes use of ANSI Z39.50, but also specifies where necessary those characteristics of GILS that are not within the scope of ANSI Z39.50. The GILS Profile does not limit how information is maintained at the source nor how the information is displayed to the user. Access to GILS is expected to be embedded within many different computer applications, ranging from the very simple to those that support concept searching across languages, dynamically interpret natural language, or filter search requests to sift huge amounts of information automatically. Public domain client software that supports access to GILS will be available from GPO, NTIS, and the Clearinghouse for Networked Information Discovery and Retrieval, among others.

Alternative ways to organize and present networked information are encouraged, but agencies participating in GILS will implement such alternatives in addition to supporting access by GILS direct users who employ the currently approved ANSI Z39.50 standard. For example, information organized via the OSI X.500 Directory Services standard can be made accessible also via ANSI Z39.50, thereby enhancing access capabilities. It should also be noted that a GILS direct user will typically use client software that provides access to a variety of information sources that do not comply with the GILS profile but are compliant with various other standards.

Some internal redundancy in GILS is to be expected--there will often be multiple GILS locator records describing the same resource and different search strategies applied by different intermediaries. Such redundancy is appropriate because the same information resources may be described differently to different audiences or for different purposes, and descriptions will cover information resources at a wide range of aggregation. Also, the same information resources may be described differently by different information services that participate directly or as intermediaries in providing Federal information to the public. Because GILS incorporates a

variety of automated and manual search techniques, questions will be answered from different perspectives depending on how GILS is used.

GPO (and perhaps NARA, NTIS and other agencies) will maintain a publicly accessible GILS source that provides a comprehensive directory of all GILS Core locator records from a Federal perspective. When appropriate to their respective missions, Federal agencies may also develop and maintain additional interagency topical locators that enhance opportunities for sharing information resources. The following are examples of topics that might be the subject of additional interagency locators: economic indicators, trade information, spatial data, educational and training resources, disaster relief, health information, biodiversity and global change research. Such locators would be similar in function to the GILS Core, but would not necessarily use the GILS Core Elements format nor be focused solely on Federal agency holdings.

GILS supports seamless access not only among locators but directly to referenced information resources. When implemented at both the client and the server, GILS linkages facilitate the electronic delivery of off-the-shelf information products, as well as connection to data systems that support analysis and synthesis of information (Figure 3). Although the trend is clearly in the direction of electronic network availability, much of the referenced information is not available currently in electronic form. GILS always provides information regarding request and delivery procedures for various distribution options as defined by the disseminating organization.

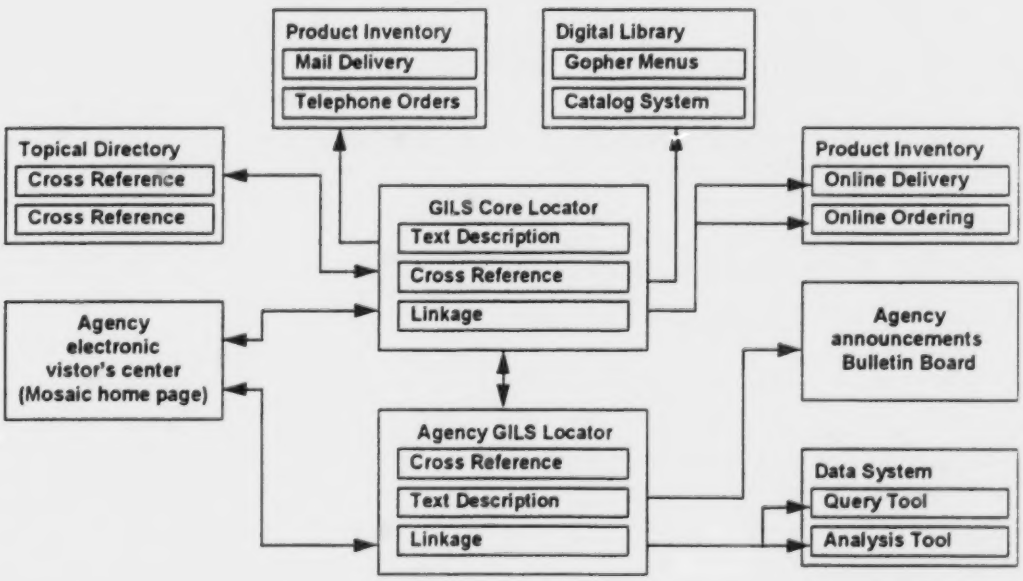

Figure 3. GILS facilitates seamless access among locators and directly to information resources.

The GILS Core

Among the GILS agency components is a set of locator records that reside on GILS accessible servers and are further identified by agencies as belonging to the GILS Core. GILS Core locator

records are required to be maintained by Federal agencies having significant information holdings, where each record describes part of the agency holdings. These Core locator records will be accessible comprehensively in the GPO Access system, but can also be aggregated by direct users of GILS to provide selective views of Federal government holdings.

The GILS Core is defined as the set of locator records maintained by the U.S. Federal government, all of which comply with the defined GILS Core Element standards, and all of which are mutually accessible through interconnected electronic network facilities. Each information disseminating agency is responsible for compiling and maintaining its own records in the GILS Core. Information services for access to GILS Core locators, once a direct user has Internet access, will be maintained by Federal agencies without charge to the direct user.

The GILS Core will include records for all information locators that catalog other publicly accessible information resources at least partially funded by the Federal government, as well as for each of the Federal government information systems that include publicly accessible data or information. While GILS Core records can point to any kind of information source, they are especially designed for helping users navigate among a wide array of other locators in various formats. It is not recommended that agencies use the precise format of the GILS Core locator records to describe all types of information resources. For example, the GILS Core Elements format would be a poor choice for describing each agency expert, but it could well be used to describe the resource that contains a compilation of such descriptions. Rather, the agency should maintain various locator records in formats appropriate to the primary user communities being served. When such other locators are published, the originating agency should include corresponding locator records that enable electronic linkage from and to the GILS Core locator.

The entire GILS Core is not likely to contain more than 100,000 locator records. In addition to locator records for information systems, it is estimated that the GILS Core will contain up to 1,000 locator records for each Federal agency that is a major disseminator of public information. Agencies that are not major disseminators will typically have fewer records in their portion of the GILS Core, especially if the agency is relatively small. Where agencies maintain information inventories that have far more records, the agency is expected to aggregate related information resources in a locator record included in the GILS Core and link the detailed inventory to GILS. Each GILS Core locator record is estimated to be less than 1,000 words in length, exclusive of any agency supplemental information that may be introduced as a separate field at the agency's discretion.

It is important to note that the vast majority of information sources accessible to GILS direct users would not be considered part of the GILS Core. Many are not maintained by the Federal Government, do not offer records in the format of the GILS Core Elements, are not on public networks, or are not offered free of charge. Many of these non-Core sources are locators nonetheless and will be very valuable to users in finding information. Also, other relevant sources of Federal information and Federal government information systems may be accessible to direct users of GILS. For example, various agencies and private-sector information providers may develop products that contain GILS Core locator records. Indeed, such derivative and value-added products may often be the first point of access to Federal information resources.

The GILS Profile

The decentralized approach envisioned for GILS requires that many different implementations be fully interoperable when implemented, although developed separately. To assure interoperability, implementors of information systems must have a clear statement of the functions of GILS and the environment within which GILS will be used. That statement becomes part of a GILS Profile that documents the specific agreements established by consensus among active implementors together with Federal representatives. The GILS Profile identifies specific standards, and the chosen subsets, options, and parameters of those standards, needed to achieve interoperability in the specific limited context of GILS.

As an initial step toward a Stable Implementors Agreement recognized by the OIW, a draft profile was created through a Cooperative Agreement between the U.S. Geological Survey and Syracuse University, with active involvement from several ANSI Z39.50 implementors representing non-government sectors. The draft GILS Profile specifies that the GILS locator records are to be available in three record syntaxes--Generic Record Syntax, United States Machine Readable Cataloging (USMARC)[7], and Simple Unstructured Text Record Syntax (SUTRS).

When using the Generic Record Syntax, the GILS locator elements can support representation in Hypertext Markup Language (HTML). (HTML is the format interpreted by the National Center for Supercomputing Applications Mosaic client software when presenting World Wide Web objects, for example.) Provision has also been made in the GILS profile to support switching among navigation techniques, including use of a browsing mode as in Gopher or a searching mode as in bibliographic systems or Wide Area Information Servers (WAIS). The incorporation in GILS of Uniform Resource Identifiers (URIs) greatly simplifies electronic navigation among locators and other data systems available on interconnected networks.

Content definitions describe the GILS Core Elements required for users to determine the relevance of defined information resources to their needs and to understand subsequent actions to obtain the information resources (see Appendix A). These definitions identify relations among GILS Core Elements, and between GILS Core Elements and the USMARC format for bibliographic data. ANSI Z39.50 definitions of GILS Core Elements in the GILS Profile provide a structure and format for movement of the GILS Core Elements between computer systems. The Abstract Record Syntax and Basic Encoding Rules used to define GILS Core Elements are also suitable for movement of element contents between automated systems using digital media such as tape, diskette, or CD-ROM.

The GILS Profile offers a preferred display format for use in printed media as well as in electronic presentations. Although specified for human viewing in English, it is intended to be extensible to other languages also.

---

[7] USMARC is an implementation of ANSI Z39.2. American National Standards Institute. (1985). American National Standard Z39.2-1985 Bibliographic Information Interchange. New York, NY: American National Standards Institute. See also USMARC Format for Bibliographic Data. Washington, DC: Cataloging Distribution Service, Library of Congress.

Appendix A. GILS Core Elements

Title: This mandatory element occurs once per locator record. It conveys the most significant aspects of the referenced resource and is intended for initial presentation to users independently of other elements. It should provide sufficient information to allow users to make an initial decision on likely relevance. It should convey the most significant information available, including the general topic area, as well as a specific reference to the subject. (USMARC Tag 245$a)

Control Identifier: This mandatory element occurs once per locator record. It is defined by the information provider and is used to distinguish this locator record from all other GILS Core locator records. The control identifier should be distinguished with the record source agency acronym as provided in the U.S. Government Manual. (USMARC Tag 001)

Abstract: This mandatory element occurs once per locator record. It presents a narrative description of the information resource. This narrative should provide enough general information to allow the user to determine if the information resource has sufficient potential to warrant contacting the provider for further information. The abstract should not exceed 500 words in length. (USMARC Tag 520)

Purpose: This mandatory element occurs once per locator record. It describes why the information resource is offered and identifies other programs, projects, and legislative actions wholly or partially responsible for the establishment or continued delivery of this information resource. It may include the origin and lineage of the information resource, and related information resources. (USMARC Tag 500)

Originator: This mandatory element occurs once per locator record. It identifies the information resource originator, named as in the U.S. Government Manual where applicable. (USMARC Tag 710$a)

Access Constraints: This mandatory element occurs once per locator record, although in some cases this element may contain the value "None." It describes any constraints or legal prerequisites for accessing the information resource or its component products or services. This includes any access constraints applied to assure the protection of privacy or intellectual property, and any other special restrictions or limitations on obtaining the information resource. Guidance on obtaining any users' manuals or other aids needed for the public to reasonably access the information resource must also be included here. (USMARC Tag 506)

Use Constraints: This mandatory element occurs once per locator record, although in some cases this element may contain the value "None." It describes any constraints or legal prerequisites for using the information resource or its component products or services. This includes any use constraints applied to assure the protection of privacy or intellectual property and any other special restrictions or limitations on using the information resource. (USMARC Tag 540)

Availability: This mandatory element occurs one or more times per locator record. It is a grouping of sub-elements that together describe how the information resource is made available.

Distributor: This mandatory sub-element occurs once per Availability element. It identifies the distributor by name, organization, street address, city, state, zip code, country, network address, hours of service, telephone, and/or fax number. (USMARC Tag 037$b)

Resource Description: This optional sub-element occurs nor more than once per Availability element. It identifies the resource as it is known to the distributor. (USMARC Tag 037$f)

Order Process: This mandatory sub-element occurs once per Availability element. It provides information on how to obtain the information resource from this distributor, including any fees associated with acquisition of the product or use of the service, order options (e.g., available in print or digital forms, PC or Macintosh versions), order methods, payment alternatives, and delivery methods. (USMARC Tag 037$c)

Technical Prerequisites: This optional sub-element occurs no more than once per Availability element. It describes any technical prerequisites for use of the information resource as made available by this distributor. (USMARC Tag 538)

Available Time Period: This optional sub-element may occur multiple times per Availability element. It provides the time period reference for the information resource as made available by this distributor. (Time period formats are as given for the Time Period of Content element described below.)

Available Linkage: This optional sub-element occurs no more than once per Availability element. It provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information. (USMARC Tag 856$u)

Available Linkage Type: This optional sub-element occurs if there is an Available Linkage described. It provides the data content type (i.e., MIME) for the referenced URI. (USMARC Tag 856 first indicator/ 856$2)

Point of Contact for further information: This mandatory element occurs once per locator record. It identifies an organization, and a person where appropriate, serving as the point of contact plus methods that may be used to make contact. Defined sub-elements include name, organization, street address, city, state, zip code, country, network address, hours of service, telephone, and fax number. (USMARC Tag 856$m for electronic resources, 535 for non-electronic resources)

Record Source: This mandatory element occurs once per locator record. It identifies the organization, as named in the U.S. Government Manual, that created or last modified this locator record. (USMARC Tag 040)

Date Last Modified: This mandatory element occurs once per locator record. It identifies the latest date on which this locator record was created or modified. (USMARC Tag 005)

Agency Program: This element occurs no more than once per locator record. It identifies the major agency program or mission supported by the system and should include a citation for any specific legislative authorities associated with this information resource. This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system. (USMARC Tag 500)

Sources of Data: This element occurs no more than once per locator record. It identifies the primary sources or providers of data to the system, whether within or outside the agency. This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system. (USMARC Tag 500)

Controlled Vocabulary: This optional element may occur multiple times per locator record. It is a grouping of sub-elements that together provide any controlled vocabulary used to describe the resource and the source of that controlled vocabulary.

> **Index Terms - Controlled:** This sub-element occurs once per Controlled Vocabulary element. It is a grouping of descriptive terms drawn from a controlled vocabulary source to aid users in locating entries of potential interest. Each term is provided in the subordinate repeating field, Controlled Term. (USMARC Tag 650)
>
> **Thesaurus:** This sub-element occurs once per Controlled Vocabulary element. It provides the reference to a formally registered thesaurus or similar authoritative source of the controlled index terms. (USMARC Tag 650 first indicator/ 650$2) Notes on how to obtain electronic access to or copies of the referenced source should be provided, possibly through a Cross Reference to another locator record that more fully describes the standard and its potential application to locating GILS information.

Local Subject Index: This optional element occurs no more than once per locator record. It is a grouping of descriptive terms to aid users in locating resources of potential interest, but the terms are not drawn from a formally registered controlled vocabulary source. Each term is provided in the repeating sub-element, Local Subject Term. (USMARC Tag 653$a)

Methodology: This optional element occurs no more than once per locator record. It identifies any specialized tools, techniques, or methodology used to produce this information resource. The validity, degree of reliability, and any known possibility of errors should also be described. (USMARC Tag 567)

Spatial Reference: This optional element occurs no more than once per locator record and provides the geographic reference for the information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."

> **Bounding Rectangle:** This optional sub-element occurs no more than once within a Spatial Reference element. It provides the limits of coverage expressed by latitude and longitude values in the order: western-most, eastern-most, northern-most, southern-most.
> (USMARC Tags 255$c, 034$d, 034$e, 034$f, 034$g)
>
> **Geographic Name:** This optional sub-element may occur multiple times within a Spatial Reference element. It identifies significant areas and/or places within the coverage through two associated constructs: a Geographic Keyword Name (USMARC Tag 651) and a Geographic Keyword Type (USMARC Tag 655). A preferred source of the names and types is the Geographic Names Information System.

Time Period of Content: This optional element may occur multiple times per locator record. It provides time frames associated with the information resource, in one of two forms:

> **Time period - structured:** Time described using the USMARC prescribed structure. (USMARC Tag 045$c)
>
> **Time period - textual:** Time described textually. (USMARC Tag 513)

Cross Reference: This optional element may occur multiple times per locator record. Each instance is a grouping of sub-elements that together identify another locator record likely to be of interest.

> **Cross Reference Title:** This optional sub-element occurs no more than once per Cross Reference element. It provides a human readable textual description of the cross reference. (USMARC Tag 787$t)
>
> **Cross Reference Linkage:** This optional sub-element occurs no more than once per Cross Reference element. It provides the machine readable information needed to perform the access (i.e., URI). (USMARC Tag 787$w)
>
> **Cross Reference Type:** This optional sub-element occurs if there is a Cross Reference Linkage described. It provides the data content type (i.e., MIME) for the referenced URI. (USMARC Tag 856 first indicator/ 856$2)

Original Control Identifier: This optional element occurs no more than once per locator record. It is used by the record source to refer to another GILS locator record from which this locator record was derived. (USMARC Tag 035)

Supplemental Information: This optional element occurs no more than once per locator record. Through this element, the record source may associate other descriptive information with the GILS Core locator record. (USMARC Tag 500)

Appendix B: Glossary

agency - any executive department, military department, government corporation, government controlled corporation, or other establishment in the executive branch of the United States Federal government, or any independent regulatory agency (OMB Circular A-130).

ANSI Z39.50 - The "American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection" is developed by the National Information Standards Organization (NISO), accredited to the American National Standards Institute (ANSI). ANSI Z39.50 complies with the Open Systems Interconnection (OSI) family of standards promulgated by the International Organization for Standardization (ISO), and is interoperable with the international standards for information search and retrieval, ISO 10162 and 10163. As of this writing, the currently approved version is ANSI Z39.50 Version 2.

direct user - a person or automated process that accesses GILS from networks using the GILS Profile and thereby having more flexibility to explore the full complement of available information. People who are direct users of GILS are assumed to be literate in English to at least the secondary school level, capable of using a personal computer, and aware of any constraints of their own hardware or software environment.

dissemination - the government initiated distribution of information to the public, excluding distribution limited to government employees or agency contractors or grantees, intra-agency or inter-agency use or sharing of government information, and responses to requests for agency records under the Freedom of Information Act (5 U.S.C. 552) or Privacy Act. Here, "disseminating information" is not distinguished from "providing access to information" (following OMB Circular A-130).

electronic information resource - information resources that are maintained in electronic, digital format and may be accessed, searched, or retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM).

government information - information created, collected, processed, disseminated, or disposed of by or for the Federal government (OMB Circular A-130).

Government Information Locator Service (GILS) - a decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

GILS Core - a subset of all GILS Locator Records which describe information resources maintained by the U.S. Federal government, comply with the defined GILS Core Elements and are mutually accessible through interconnected electronic network facilities without charge to the direct user.

government publication - information that is published as an individual document at government expense, or as required by law (OMB Circular A-130).

information - any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms (OMB Circular A-130).

information product- any book, paper, map, machine-readable material, audiovisual production, or other documentary material, regardless of physical form or characteristic (OMB Circular A-130).

information resource - includes both government information and information technology (OMB Circular A-130).

information service - considered equivalent to information product from the policy perspective of OMB Circular A-130, although agency locator records for services may differ from those for products.

information system - the organized collection, processing, maintenance, transmission, and dissemination of information in accordance with defined procedures, whether automated or manual (OMB Circular A-130).

information technology - the hardware and software operated by a Federal agency or by a contractor of a Federal agency or other organization that processes information on behalf of the Federal Government to accomplish a Federal function (OMB Circular A-130).

intermediary or intermediate service - an entity or service that makes some of the GILS information available but does not provide the full capabilities of a direct user.

interoperability - a condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

locator - an information resource that identifies other information resources, describes the information available in those resources, and provides assistance in how to obtain the information.

locator record - a collection of related data elements describing an information resource, the information available in the resource, and how to obtain the information.

mandatory element - a data element in a GILS Core Locator Record that must have a value provided by the record source.

Open Systems Interconnection (OSI) - a family of standards promulgated by the International Organization for Standardization (ISO) and adhering to a specific model that promotes interoperability.

profile - the statement of a function(s) and the environment within which it is used, in terms of a set of one or more standards, and where applicable, identification of chosen classes, subsets, options, and parameters of those standards; a set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

records management - the planning, controlling, directing, organizing, training, promoting, and other managerial activities involved with respect to records creation, records maintenance and use, and records disposition in order to achieve adequate and proper documentation of the policies and transactions of the Federal government and effective and economical management of agency operations. (44 U.S.C. 2901(2))

Uniform Resource Identifier (URI) - a set of related standards for encoding resource location and identification information for electronic and other objects. Examples include Uniform Resource Locators (URLs) and Uniform Resource Names (URNs).

USMARC - an implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data. This documentation is published by the Library of Congress.

Attachment F

EXPANDING RESEARCH AND DEVELOPMENT ON THE ANSI/NISO Z39.50 SEARCH AND RETRIEVAL STANDARD

PROJECT WORKSTATEMENT

Introduction

This project workstatement accompanies the contract for services each of the research team members will sign. Subsequent sections of this workstatement detail general responsibilities of the core group and also describes specific tasks that members of the research team will carry out.

The research team for this project includes a core group of knowledgeable and experienced people who will meet together three times and will be primarily responsible for carrying out Project Tasks 1-4. The research team also includes additional consultants who are responsible for Tasks 5-7. The latter group will not necessarily attend the meetings of the core group.

Intellectual Property and Copyrights

As a Federally funded and sponsored project, all products of this project, including electronic documents and documented computer code, will be placed in the public domain and will be released without any restrictions on future use. As original authors, some of the consultants participating in this project may hold rights to extant computer code and documentation components that may be incorporated into the products of this product. These individuals and related organizations holding such rights will be requested to forgo their rights to restrict any future use of those components.

The co-principal investigators retain rights of first refusal for any and all consultant products, in whole or in part, for inclusion into the final report that will be developed as part of this project. Consultants are encouraged to use and disseminate their products for this project after the final report has been submitted.

General Responsibilities of Core Group Members

The project's success will depend on the participation of core group members in the research team. Core group members will have responsibilities to:

- Attend three meetings in Washington, DC
- Assist in determining project plan, timelines, and activities
- Assume project-related staff responsibilities, (e.g., editor, recorder, librarian, etc.)
- Act as a nominal representative of community(ies) with which you have connection or affliation
- Be committed to seeing the project through to completion (within the general timeframe specified in the contract)
- Assist in building consensus of team members and of identified stakeholders on application profile
- Identify outside experts as needed
- Comment on draft and final documents by other team members
- Produce recommendations for ZIG and for further research, development, etc. of profile.

Fulfilling these responsibilities will enable the team to work effectively and complete its work in a timely manner.

Specific Tasks for Research Team

This section details the individual tasks that will enable the project to accomplish its goal and objectives. Each task statement includes the following components:

- Description of task
- Goal of task
- Objectives of task
- Process by which task is accomplished
- Deliverable or end product of each task.

Tasks 1-4 are core group members' tasks:

Task #1: Stakeholder Contact for Consensus Building
Task #2: Profile Areas for Research and Developmen
Task #3: Document How Profile Specifically Addresses (or doesn't) Stakeholder Concerns
Task #4: Develop Test Scenarios and Validation Activities for Test Implementations.

The remainder of the tasks will be carried out by other members of the research team who may not be part of the core group. These tasks are:

Task #5: Feasibility and Requirements for Interoperability Testing and Conformance Testing
Task #6: Evaluation of Research Project Documents and Activities
Task #7: Summary Report of Project and Document Compilation.
Task #8: Critical Review of WAIS as a Tool for GILS
Task #9: Profile Requirements for Accommodating Information Systems Information and Records Management Needs

Task 1: Stakeholder Contact for Consensus Building

Description: Team members will be assigned particular stakeholders or stakeholder communities to contact for input into profile development activities. Team members will jointly identify potential contacts, and individual team members will initiate contact to determine stakeholder interests and concerns. Based on the initial contacts, project coordinators will mail out an information packet to the stakeholder. Upon receipt of the packet and time to read it, the individual team member responsible for that contact will do a follow-up interview to gather any specific or general concerns and issues the stakeholder identified. Project coordinators will supply a "team developed" interview protocol and form for data collection for each contact and follow-up. Individual team members will be responsible for completing a data collection form and submitting it to project coordinators for compilation and consolidation. Team members should anticipate approximately 6 contacts each.

Goal: Bring awareness to potential stakeholder communities of the profile development and build consensus among stakeholder communities by providing them input into the profile development process.

Objectives: 1) Inform potential stakeholder about the project by making an initial contact with stakeholder.

2) Solicit input from stakeholders about their concerns and issues based on a knowledge of this profile development project.

3) Address concerns and issues by collecting data and using the data as a basis for profile considerations.

Process: Telephone calls or electronic mail messages for the initial contact. Offer to send an information packet to potential stakeholder. Follow-up with a telephone call to each stakeholder who received a packet within a week after the packet is mailed from the project coordinators. Document the responses from the stakeholder on a prescribed form and submit the completed form to the project coordinators.

Product: 1) List of contacts made.

2) Completed data collection forms submitted to project coordinators within 5 working days after follow-up conversation.

Person: Core group members.

Task 2: Profile Areas for Research and Development

Description: Given the draft outline of the profile, there are a number of areas that must be addressed in more complete form. Some of these profile areas will be directly related to Z39.50 and others will be unrelated to Z39.50. Individual team members will be assigned to one or more profile areas (e.g., attribute set, record syntax). Each area will need to be examined, researched, and addressed by a team member. The team member is responsible for developing this area of the profile.

Goal: To produce profile specifications for each area of the team developed draft profile outline.

Objectives: 1) Research the specific area of the profile to be developed.

2) Develop an appropriate specification for each area.

Process: Individual team members will work independently on their assigned areas. They should rely on their own expertise, contact with other experts, and discussions with other team members in developing specifications for the profile. After developing the specifications, team members will submit their documents to the project coordinators.

Product: 1) Written specifications for each assigned area.

2) Other information (e.g., written background information, other materials, etc.) upon which the specifications are based.

Person: Core group members.

Task 3: Document How Profile Specifically Addresses Stakeholder Concerns

Description: Since an important component of this project is consensus building of stakeholders, individual team members will document how, in their assigned areas of profile development, they dealt with and responded to specific stakeholders' concerns.

Goal: To build consensus among all stakeholders by addressing their concerns in the development of the profile.

Objectives: 1) Acknowledge the concerns of the stakeholders in the specific areas of profile development for which the individual team members are responsible.

BEST COPY AVAILABLE

2) Document specifically how the profile addresses (or doesn't) the concerns of the stakeholders.

Process: Individual team members will receive the stakeholder concerns for the area(s) in which they are developing profile specifications. They should take account of these stakeholder concerns when developing the profile. In addition, they will document in written form how each stakeholder concern was addressed by the profile.

Product: A written document that summarizes stakeholder concerns for specific area(s) of profile, presents the details on how the individual team member dealt with stakeholder concerns in the profile. If the concern was not able to be addressed in the profile, this should be noted. Provide recommendations for future action to address stakeholder concerns.

Person: Core group members.

Team Task 4: Develop Test Scenarios and Validation Activities for Test Implementations

Description: To evaluate adequately the test implementations that groups will develop based on the "draft profile," a series of test scenarios and validation activities are needed. Based on previous experience with interoperability testing, individual team members will develop scenarios to validate the test implementations' compliance with the profile and the interoperability among the test implementations. The profile and its relationship to the functional and technical requirements provided to the team will be tied directly to test scenarios. In addition, guidelines for testing the implementations will be developed. This may also serve as the foundation of interoperability testing and conformance testing requirements.

Goal: To establish the interoperability of the test implementations developed by a number of implementors.

Objectives: 1) Develop test scenarios which can be used to validate the test implementations.

2) Identify ways in which to evaluate the test scenario results.

3) Determine validation techniques.

Process: Team members will identify a set of test scenarios and document how these scenarios are tied to the profile specifications and the technical and functional requirements for the profile. Guidelines for the actual testing will also be developed. A background document by a research team member will provide information and recommendations for this task (see Task #5).

Product: A series of written documents that include:

1) Test scenarios.

2) Guidelines for testing.

3) Evaluation criteria.

Person: Core group members

Task 5: Feasibility and Requirements for Interoperability Testing and Conformance Testing

Description: A primary goal of profile development is to ensure interoperability and interworking of implementations developed in accordance with the profile. Interoperability and conformance testing are complex activities, and as of now, there is little guidance available in such testing for Z39.50. This task will involve research and development of possible options for interoperability and conformance testing. While the core group members will be assist in developing test scenarios for the test implementations, this task will result in a document to guide the development of the test scenarios. In addition, it will provide a wider overview of interoperability and conformance testing for consideration by team members and members of the OIW Library Applications Special Interest Group.

Goal: Provide an understanding of the need for and limitations of interoperability and conformance testing for Z39.50, and specifically for use with the locator application profile.

Objectives:

1) Provide background information on interoperability and conformance testing in the open systems environment. This will include both Federal government interoperability and conformance testing requirements as well as such testing that is occurring under the auspices of national and international standards organizations and others.

2) Develop options for testing interoperability and conformance of the locator application profile.

3) Suggest a suite of test scenarios for the locator application profile.

4) Provide a sense of the feasibility of interoperability and conformance testing for Z39.50 applications and implementations (e.g., the usefulness of the Z39.50 Implementors Testbed [ZIT]).

Process: To be determined by the consultant, but should include information gathering from authoritative sources for interoperability and conformance testing.

Product: A written document that summarizes issues related to feasibility of interoperability and conformance testing for Z39.50 applications. It presents options for testing interoperability and conformance testing. In addition, the document will recommend a suite of test scenarios for the locator application profile for the test implementations; these may serve as the basis for more formal interoperability testing. The document will be submitted to project coordinators.

Person: Cecilia M. Preston

Task 6: Evaluation of Research Project Documents and Activities

Description: Throughout the project, team members will be producing documents related to individual task assignments and group work. To ensure an objective and ongoing review of the project, this task will have a consultant review and comment on the major products of the project. This task will include a summary review of the accomplishments of the projects and recommendations for further action regarding this profile and project.

Goal: Ensure that the project is subject to third-party review and assessment by a knowledgeable and experienced person.

Objective: 1) Provide review and comments on working documents and final products.

2) Suggest mid-course corrections in project activities and directions, if necessary.

3) Provide an objective, neutral perspective on the work of the team.

Process: Review over the course of the project the relevant working and final documents and provide comments to project coordinators. Person will be available for consultation in face-to-face or telephone conversations, or other means.

Product: Written and verbal responses to the work of the research team and the project. These responses should address the objectives of this task by summarizing the consultants review of the documents and major products of the project. In addition, the response should include recommendations for modification of project activities or specific products. A short (approximately five pages) summary document will assess the overall accomplishments and shortfalls of the project and will include recommendations for further action and next steps regarding the profile, the project, or in other areas.

Person: Clifford A. Lynch

Task 7: Summary Report of Project and Document Compilation

Description: As a final task for the project, this will involve the writing of a summary report of the activities and results of the project. It will include the editing and compilation of relevant documents produced by the project. The responsibilities for this task also involves dissemination of project activities and accomplishments using various mechanisms (e.g., electronic dissemination, documents on an FTP site, etc.).

Goal: To provide summary documentation of the entire research project for distribution to stakeholders.

Objectives: 1) Create a written summary of the project, its activities, and accomplishments.

2) Edit, revise, and compile pertinent documents and reports produced in the course of the project.

3) Prepare the final report for submission to project sponsors that will include pertinent project documents (edited, revised, and compiled) and the summary of the project, its activities, and accomplishments.

4) Disseminate information about the project.

Process: Consultant will write a summary based on the work of the research team. This will include editing, revising, compiling, and consolidating the relevant documents documents produced during the course of the project. A final report will be completed consisting of the summary and edited project documents. The final report will be submitted to project coordinators. The consultant will also have responsibility for suggesting and carrying out dissemination activities.

Product: A written document consisting of the summary of project activities and accomplishments accompanied by edited and consolidated documents produced during the project documentation. The contents and format of the final report will depend on the nature of the reports produced by the research team. This final report will be the basis of information for dissemination.

Person: William E. Moen

Task 8: Critical Review of WAIS as a Tool for GILS

Description: This task involves a critical review of Wide Area Information Servers (WAIS) as an application tool for the Government Information Locator Service (GILS). The consultant will evaluate the effectiveness of the current WAIS implementation based on Z39.50 (i.e., the 1988 version of the standard) to achieve the functionality of the proposed GILS as described in the Government Information Locator Service overview document (November 11, 1993 Draft) and the document developed by the research team, Using Z39.50 in an Application for the Government Information Locator Service: A Background Paper (January 1, 1994 Draft). In addition, the consultant will evaluate the proposed implementations of WAIS as defined in the WAIS Profile (Draft) which will be an implementation of WAIS that is aligned with Z39.50 — 1992. The consultant will identify strengths and weaknesses of WAIS (under Z39.50 — 1988 and 1992) in providing the functionality required for GILS. The consultant will prepare a technical report covering background on WAIS and the findings of this critical review. The report will include recommendations for changes to the GILS Profile based on the review and/or suggest changes to the WAIS profile to address compatibility with GILS requirements.

Goal: Provide the research project with an increased understanding of current WAIS functionality vis a vis GILS requirements as well as how proposed implementations of WAIS based on Z39.50 — 1992 will respond to GILS requirements.

Objectives:

1) Review and evaluate WAIS (Z39.50 — 1988) and the WAIS profile for Z39.50 — 1992 with respect GILS requirements.

2) Identify weaknesses and strengths of WAIS (under Z39.50 — 1988 and 1992) in achieving the functionality required by GILS.

3) Make recommendations and/or offer cautions regarding the use of WAIS (under Z39.50 — 1988 and 1992) in GILS applications.

4) Suggest changes to GILS Profile based on functionality of WAIS and/or changes to WAIS profile to address compatibility with GILS requirements.

Process: Consultant will review requirements of GILS as specified in two documents: Government Information Locator Service overview document (November 11, 1993 Draft) and Using Z39.50 in an Application for the Government Information Locator Service: A Background Paper (January 1, 1994 Draft). Consultant will review WAIS as currently implemented (i.e., based on Z39.50 — 1988) and as proposed in the WAIS Profile for Z39.50 — 1992. Consultant will evaluate the suitability of WAIS as a tool for the GILS. Consultant will, if necessary or requested by Project Coordinator, talk to parties involved in defining the GILS requirements and the GILS profile. Consultant will prepare and submit a technical report on the findings of the critical review. A draft of the report will be submitted by the consultant to the Project Coordinator on or about January 15, 1993 for review and comment. A final report, acknowledging and incorporating where necessary the Project Coordinator's comments, will be submitted by the contractor to the Project Coordinator by January 25, 1993.

Product: A technical report containing: 1) an overview and background of functionality of WAIS for use in the GILS application; 2) a detailed and critical listing of how WAIS (as implemented under Z39.50 — 1988 and as proposed under Z39.50 — 1992) does and/or does not address the functionality required by GILS. The report should identify both the strengths and weaknesses of WAIS (implemented and proposed) as a tool for GILS. Consultant should provide recommendations and cautions on the use of WAIS for GILS. Consultant will include specific recommen-

dations, if any, for changes to the GILS Profile based on the functionality offered by WAIS and/or include suggested changes in the WAIS profile to address compatibility concerns with GILS requirements. The report should include necessary appendices and attachments as appropriate to facilitate understanding of the report.

Person: Francois Schiettecatte, FS Consulting

Task 9: Profile Requirements for Accommodating Information Systems Information and Records Management Needs

Description: This task involves the development of a background paper outlining the needs of potential user communities of the Government Information Locator Service (GILS). Among the many Federal information resources that are in the category of public information are particular information systems that house publicly accessible records of the business and operations of Federal agencies. The archival and records management user communities are interested in having the GILS Profile support adequate description of and access to these resources. In addition to outlining the needs of these user communities, this paper will describe any enhancements and changes to the functional and service requirements of the GILS as documented in the Government Ifnormation Locator Service (GILS) (current draft dated January 19, 1994), and to the draft specifications for the GILS Profile as documented in "Using Z39.50 in an Application for the Government Information Locator Service (GILS)" (current draft dated January 24, 1994).

Goal: Provide the research project with an increased understanding of the needs of the archival and records management communities regarding description of and access to Federal information resources and suggest enhancements or changes to the developing GILS Profile to support the requirements based on these needs.

Objectives:

1) Outline the scope of the problem for identifying, describing, and accessing information about Federal information resources that are of interest to the archival and records management communities.

2) Describe the unique information needs of these communities.

3) Identify the functional requirements that can be derived from these information needs.

4) Suggest enhancements or changes to the emerging GILS Profile to support these functional requirements.

Process: Consultant will prepare a written report that responds to the objectives of this task. Consultant will review the current drafts of the two primary documents (listed above) that are the basis for the emerging GILS Profile, and based on this review will compare the information needs of the specific user communities (i.e., archival and records management) and the ability of the proposed GILS and the emerging GILS Profile to respond to those needs. Consultant will identify specific recommendations for enhancements and/or changes to the GILS Profile that will increase the functionality of the GILS to support the description of and access to the Federal information resources of concern to these communities.

A draft of the report will be submitted to the Project Coordinator on or about February 20, 1994 for review and comment. A final report, acknowledging and incorporating where necessary the Project Coordinator's comments, will be submitted by the contractor to the Project Coordinator by February 28, 1994.

Product: A report containing: 1) an outline of the scope of the problem of two specific user communities in identifying, describing, and accessing Federal information resources; 2) a listing of the information needs of these communities and the descriptions of functional requirements that derive from these needs that the GILS should support; 3) specific recommendations for enhancements or changes to the emerging GILS Profile that addresses these functional requirements. The report should include necessary appendices and attachments as appropriate to facilitate understanding of the report.

Person: David Bearman

Attachment G

MEETINGS OF THE GILS PROJECT TEAM

1) Z39.50 Implementors Group Meeting, October 4, 1993, Ottawa, Canada

   - Introductions and preliminary discussions with potential project team members.

2) GILS Project Team Meeting, October 28, 1993, Washington, DC

   - First working meeting of team.

3) GILS Project Team Meeting, December 4-5, 1993, Washington, DC

   - Second working meeting of team. Preceded an OIW meeting at which an update on the work of the GILS was presented.

4) GILS Project Team Meeting, January 14-15, 1994, Washington, DC

   - Third and final official working meeting of the team.

5) Z39.50 Implementors Group Meeting, January 26-28, 1994, Gainesville, FL

   - Presentation to meeting on GILS Profile.

6) Open System Environment Implementors' Workshop Meeting, March 15, 1994, Washington, DC

   - Half day discussion of GILS Profile to prepare for its submission to OIW.

7) Interim OIW Meeting and Z39.50 Implementors Group Meeting, April 26-28, 1994

   - OIW meeting to make final revisions to the GILS Profile.

Attachment H

STAKEHOLDER CONTACT LIST

This list includes the names of people and organization the GILS project leaders or project team members contacted or were contacted by in the course of developing the GILS Profile. The contacts were made through letter mail, electronic mail, telephone, or in person.

Prudence Adler
Association of Research Libraries
Washington, District of Columbia

Scott Armstrong
Information Trust
Washington, District of Columbia

David Bearman
Achives and Museum Informatics
Pittsburgh, Pennsylvania

David R. Bender
Special Libraries Association
Washington, District of Columbia

Eric Bivona
Dartmouth College
Hanover, NH

Julia C. Blixrud
Council on Library Resources
Washington, DC

Charles R. Blunt
State University of New York
Albany, NY

Cheryl Boettcher
University of California, Los Angeles
Los Angeles, CA

Marilyn Cade
AT&T
Washington, DC

Dan Cantrall
Oregon State Archives
Portland, OR

Priscilla L. Caplan
University of Chicago Library
Chicago, IL

Bonnie C. Carroll
CENDI
Oak Ridge, TN

John Churchman
Smithsonian Institute
Washington, DC

Karen Coyle
University of California
Oakland, CA

David S. Day
MITRE Corporation
Bedford, MA

Vincent DeSanti
Project Management Enterprises
Bethesda, MD

Kathleen M. Eisenbeis
Universities Space Research Association
Washington, District of Columbia

Miles Fidelman
Center for Civic Networking
Charlestown, MA

Maggie Freed
University of Southern California
Los Angeles, CA

Patrick Frisbie
NOTIS Systems
Evanston, IL

John R. Garrett
Corporation for National Research Initiatives
Reston, Virginia

Jonathon P. Gill
White House Office of Media Affairs
Washington, District of Columbia

Myke Gluck
Florida State University
Tallahasee, FL

Richard Griffin
Oregon State University
Corvallis, OR

Howard Harris
RMG Consultants
Bethesda, Maryland

Patricia Harris
National Information Standards Organization
Gaithersburg, MD

Carol C. Henderson
Washington Office American Library Association
Washington, District of Columbia

Linda L. Hill
NASA
Washington, DC

Diane I. Hillman
Cornell Law Library
Ithaca, NY

Peter Bolles Hirtle
National Archives and Records Administrion
Washington, DC

Bernadine Abbott Hoduski
Joint Committee on Printing
Washington, District of Columbia

Timothy C. Judkins
UT Southwestern Medical Center
Dallas, TX

Brian Kahin
John F. Kennedy School of Government,
Science, Technology and Public Policy
Cambridge, Massachusetts

Mark Kaprow
General Services Administration
Washington, DC

Neal K. Kaske
United States Department of Education
Washington, District of Columbia

Anna D. Keller
Library of Congress
Washington, DC

John Kunze
University of California, Berkeley
Berkeley, CA

Mary Levering
Federal Library and Information Center Committee
Library of Congress
Washington, District of Columbia

David Loy
Dialog Information Services
Sunnyvale, CA

John Mallery
Massachusetts Institute of Technology
Cambridge, MA

Sally H. McCallum
Library of Congress
Washington, DC

Brad McLean
Gaylord Information System
Liverpool, NY

Paul Evan Peters
Coalition for Networked Information
Washington, District of Columbia

Sara Randall
NOTIS Systems
Evanston, IL

Peter Ryall
Mead Data Central
Miamisburg, OH

Winston Tabb
Library of Congress
Washington, District of Columbia

Paul Vassallo
National Institute of Standards and Technology
Gaithersburg, Maryland

Janet Vratny
Apple
Cupertino, CA

Lisa B. Weber
National Historical Publications and Records Commission
Washington, District of Columbia

Duane E. Webster
Association of Research Libraries
Washington, District of Columbia

Paul Weiss
National Library of Medicine
Bethesda, MD

Les Wibberly
Chemical Abstracts Service
Columbus, OH

Gregory Wool
Iowa State University
Ames, IA

Peter Young
National Commission on Libraries and Information Science
Washington, District of Columbia

Queinnec Young-Hee
National Library of Canada
Ottawa, Ontario

FS Consulting

Technical Report:

Critical Review of WAIS as an Application Tool for GILS

Prepared by:
F. Schiettecatte,
FS Consulting,
435 Highland Avenue,
Rochester, New York, 14620
(716) 256-2850 (Tel)
(716) 473-9695 (Fax)

Critical Review of WAIS as an Application Tool for GILS

1. Background:

This technical report is a critical review of Wide Area Information Servers (WAIS) as an application tool for the Government Information Locator Service (GILS).

The goal of this report is to identify the strengths and weaknesses of WAIS under Z39.50 -- 1988 (Z39.50, Version 1) and Z39.50 -- 1992 (Z39.50, Version 2) [1] in providing the functionality required for the proposed GILS as described in the _Government Information Locator Service overview document (January 22, 1994 Draft)_ [2], and in the profile document developed by the research team, _Using Z39.50 in an Application for the Government Information Locator Service: A Background Paper (January 24, 1994 Draft)_ [3].

This report will first evaluate the effectiveness of the current WAIS implementation based on Z39.50, Version 1, hereafter referred to as WAIS-88, in providing the functionality of the proposed GILS. A copy of the document _WAIS over Z39.50-1988_ which defines the WAIS-88 profile is included in Appendix A.

This report will then perform the same evaluation against the proposed implementations of WAIS as defined in the WAIS Profile (Draft) which will be an implementation of WAIS that is aligned with Z39.50, Version 2, hereafter referred to as WAIS-92. A copy of the document _WAIS Profile Of Z39.50-1992, Version 1.1_ which defines the WAIS-92 profile is included in Appendix B.

In addition, this report will make suggestions for changes to the WAIS-92 profile to address compatibility issues with the GILS profile, and also suggest which of the WAIS-92 profile features could incorporated into the GILS profile.

2. Scope:

The scope of this report is limited to comparing and contrasting the various profiles at a protocol level. This report will not look at Server (defined as a Target in Z39.50) or Client (defined as an Origin in Z39.50) implementation issues at all. For example no attention will be paid to the retrieval method a

Server might use to search for data, or to how a Client might prompt a user for search requests or present back various types of data to a user [4]. This is a very important point to bear in mind throughout this report because there are many components that make up any Client-Server based information system and the protocol used to exchange information is just one component of such a system. In fact, users in general do not "perceive" all the interactions that may go on behind the scenes as they use the system, where a software engineer would (justifiably) distinguish different components, the user will generally only see one, the whole system.

3. Terminology:

Various terms are used in this report. These are defined below:

"Server" defined as a Target in Z39.50 (all versions).
"Client" defined as an Origin in Z39.50 (all versions).
IETF corresponds to the Internet Engineering Task Force [5].
URL corresponds to a Universal Resource Locator as defined by the IETF [6].
URN corresponds to a Universal Resource Name as defined by the IETF [6].
Z39.59, Version 1 corresponds to Z39.50 – 1988 [1].
Z39.59, Version 2 corresponds to Z39.50 – 1992 [1].
Z39.59, Version 3 corresponds to Z39.50 – 1994 [1].

4. Comparison of the GILS profile and WAIS-88:

This section of the report will use a structure similar to that used in the background paper _Using Z39.50 in an Application for the Government Information Locator Service: A Background Paper (January 24, 1994 Draft)_. This was deliberately done so that the reader could easily compare the WAIS-88 profile review below with the corresponding requirements of the GILS profile. It should be noted here that the WAIS-88 profile has been in use for approximately four years and has not been changed in any way since it was first released to the public domain, neither is it expected to change at all in the future. Where the WAIS-88 profile does not support features required by the GILS profile, possible work-arounds will be suggested if technically feasible, to enable the WAIS-88 profile to support required features.

4.1 Assumptions and Agreements about GILS

4.1.1 GILS User Model:

The basic Locator Record requires the inclusion of a Local Control Number, a Control Identifier and an Availability Element. While the WAIS-88 profile does not support these three elements directly, they could be supported by using the Document Identifier element defined in the WAIS-88 profile. This Document Identifier element includes some of the functionality of the three elements listed above. It should be noted that in Z39.50, Version 1, the composition of the Document Identifier element is not specified and is left up to the implementor of the profile.

In the WAIS-88 profile, the Document Identifier element is an object generated by a WAIS Server which allows a Client and/or Server to uniquely identify a record. This Document Identifier element is a relatively complex object, with a number of sub-elements, and is guaranteed to be unique for any record generated by any Server. It is guaranteed to be unique because it includes the Internet host name of the machine, on which the document resides, and the full path name of the file along with some ancillary positional information. The composition of a Document Identifier element is described in more detail in the document _Document Identifiers or International Standard Book Numbers for the Electronic Age_.

The various sub-elements that make up the Document Identifier element defined in the WAIS-88 profile include the Original Server sub-element, the Original Database sub-element and the Original Local ID sub-element. These three sub-elements are replicated as Distributor fields as well, so there is a Distributor Server sub-element, a Distributor Database sub-element and a Distributor Local ID sub-element. There is also a Copyright Disposition sub-element.

The functionality required by the Local Control Number element can be mapped to the various Distributor sub-element, the Control Identifier element can be mapped to the various Original sub-element and the Availability Element can be mapped to the Copyright Disposition sub-element. The WAIS-88 profile does not specify the format and composition of the Copyright Disposition and it has never been used (to the best knowledge of this author). The mapping between GILS elements and WAIS-88 Document Identifier sub-elements could be done as follows:

GILS Element	WAIS-88 Sub-Elements
Local Control Number	Distributor Database/Distributor Local ID
Control Identifier	Original Database/Original Local ID
Availability Element	Copyright Disposition

The major drawback of this approach is that, in the WAIS-88 profile, a Client should not have any knowledge of the composition of the Document Identifier element, i.e. the Document Identifier element should be entirely "opaque" to Clients. These mappings and the requirement of the GILS profile in allowing Clients to identify duplicate records following a search would force them to have complete knowledge of the internal structure of a Document Identifier element including its' sub-elements. The only problem is that this prevents a Server from changing the structure and composition of its' Document Identifier elements in an arbitrary fashion (this has not been a common occurrence). It should be noted that the Copyright Disposition sub-element will probably not suffice for the requirements of the Availability Element as defined by the GILS profile.

It should also be noted that while this solution is technically feasible, it cannot be considered to be ideal as it would "overload" the functionality of the Document Identifier element and would use it in a way that is not its defined use.

4.1.2 GILS Searching:

The WAIS-88 profile does not support fielded searching at all and does not comply with the GILS profile in this area. Furthermore when field searching has been implemented in WAIS-88 compliant Servers, it has not been done in a Z39.50 compliant manner (see section 4.2.3. for a greater explanation of searching in the WAIS-88 profile).

4.1.3 GILS Browsing:

While there is no explicit support in the WAIS-88 profile for hierarchical browsing of menu structures, this could be implemented relatively easily by mapping the WAIS Headline record to the GILS Cross Reference record, and getting the server to return a specific record when queried using the "well-know" search (see section 4.2.10). The mapping between GILS elements and WAIS-88 headline record elements could be done as follows:

GILS Element	WAIS-88 Element
Title	Headline
URL	Document Identifier

4.2 Z39.50 Specifications for the GILS Application:

4.2.1 Version:

GILS is designed to operate on Z39.50, Version 2 and the WAIS-88 profile operates on Z39.50, Version 1. However there is no technical reason why the GILS profile could not operate on Z39.50, Version 1.

4.2.2 Facilities:

GILS requires the use of the following Z39.50, Version 2 Services; Init, Search, Present and Termination [8]. Of these, the Present Service is the only one not supported in the WAIS-88 profile. The retrieval of documents is achieved though the Search Service combined with the use of the Type-1 query. The exact mechanics of how this is performed are described in _WAIS over Z39.50-1988_ , _WAIS Interface Protocol, Prototype Functional Specification_ and _WAIS Protocol Users Manual_.

Standard Z39.50, Version 1 Init Service negotiation procedures are supported by the WAIS-88 profile.

4.2.3 Search Service Parameters:

While the GILS profile requires the support of Type-1 queries, the WAIS-88 profile uses Type-3 queries for searches. This query type is defined specifically by the WAIS-88 profile to enable clients to send a text string, entered as a search by the user, to a Server without being parsed out into Reverse Polish Notation (RPN) queries. Type-1 queries are supported in the WAIS-88 profile for document retrievals only and use their own, non-standard, attributes (see section 4.2.2).

4.2.4 Attribute Sets:

Since the WAIS-88 profile only supports a special form of the Type-1 query (for document retrievals only) and uses a Type-3 query for searches, there is no support for the Bib-1 Attribute Set. The WAIS-88 profile was also designed to operate on databases of free-text documents where the document itself was the basic record and had no formal structure in the Z39.50 sense, however the document may have a structure or be encoded in a special manner (see section 4.2.6).

4.2.5 Diagnostic Messages:

The WAIS-88 profile supports the Bib-1 diagnostic set required by Z39.50, Version 1 (even though some of the diagnostics may not make sense in the context of the WAIS-88 profile due to the nature of the profile).

4.2.6 Record Syntax:

The WAIS-88 profile does not implement a particular record syntax for the presentation of documents and supports unstructured text for the retrieval of documents. In that respect, the WAIS-88 profile does not meet the basic requirements of the GILS profile for supporting Unstructured Text (SUTRS) records, USMARC records and Generic Record Syntax (GRS).

However it should be noted that the support for unstructured text in the WAIS-88 profile is not just limited to text. Binary data such as images or complex structured records can be transferred between Server and Client, so the transfer of USMARC records or GRS records could be supported in the WAIS-88 profile. The documents requested by a Client and returned by a Server are tagged with a Document Type which defines the format of the document. Historically such types as 'TEXT', 'PICT', 'GIF', etc., have been used. One could easily implement a USMARC Document Type and/or a GRS Document Type which could be parsed by the Client in some meaningful way. Alternatively one could default to an Unstructured Text (SUTRS) record version of the Locator Record as a 'TEXT' Document Type which would be the simplest and easiest solution to implement, though this would be a bare minimum as far as the GILS profile is concerned.

4.2.7 Preferred Presentation Format:

The preferred presentation format for SUTRS records is a Server implementation issue and is not part of the protocol.

4.2.8 Schema:

Since the WAIS-88 profile does not directly support GRS records, the only way to support a specific GRS schema would be to support GRS records along with the newly defined Schema-1 schema [9] (see section 4.2.6).

4.2.9 Element Set Names:

Since the WAIS-88 profile does not support element sets, element set names are not supported.

4.2.10 The Well-Known Search:

While there is no explicit support in the WAIS-88 profile for a "well-known" search, this could be implemented relatively easily by causing the Server to return a specific document when it receives a query with a term of zero length (this is already the case in most WAIS-88 Server implementations).

4.3 Data Elements in the Locator Records:

Since the WAIS-88 profile only supports basic documents (see section 4.2.6 for greater detail), data elements are not supported. Support for data elements could be implemented in the manner described in section 4.2.6.

5. Comparison of the GILS profile and WAIS-92:

This section of the report will use a structure similar to that used in the previous section. Again this is done for ease of comparison between the WAIS-92 profile review and the corresponding requirements of the GILS profile. It should be noted here that the WAIS-92 profile is still evolving as a profile and as such is subject to changes. The WAIS-92 profile used was current at the time this report was written, but may have changed following the publication of this report. The actual WAIS-92 profile document used for the comparison is included in Appendix B. Suggestions will be made as to how the WAIS-92 profile may be brought into compliance with the GILS profile where applicable.

5.1 Assumptions and Agreements about GILS

5.1.1 GILS User Model:

The basic Locator Record requires the inclusion of a Local Control Number, a Control Identifier in the form of a URL (Uniform Resource Locator) and an Availability Element. These are partially supported in the WAIS-92 profile, the mapping between element names would be done as follows:

GILS Element	WAIS-92 Element
Local Control Number	Local Number
Control Identifiers	URx

No support exists for the Availability Element in the WAIS-92 profile and this would need to be added for conformance with the GILS profile.

5.1.2 GILS Searching:

The WAIS-92 profile fully support the Type-1 Queries and thus will support the GILS profile requirement for fielded searches.

5.1.3 GILS Browsing:

While there is no explicit support in the WAIS-92 profile for the browsing of hierarchical menu structures, this could be implemented relatively easily by mapping the WAIS record to the GILS Cross Reference record, and getting the Server to return a specific record when queried using a "well-known" search (see section 5.2.10). The mapping between GILS elements and WAIS-92 elements could be done as follows:

GILS Element	WAIS-92 Element
Title	Headline
URL	URx

This could be implemented without requiring any changes to the WAIS-92 profile.

5.2 Z39.50 Specifications for the GILS Application:

5.2.1 Version:

The WAIS-92 profile is designed to operate on Z39.50, Version 2 like the GILS profile.

5.2.2 Facilities:

The WAIS-92 profile requires full conformance with V39.50, Version 2 and will support all the Services required by the GILS profile, these include the Init, Search, Present and Termination Services. Standard Z39.50, Version 2 Init Service negotiation procedures are supported in the WAIS-92 profile.

5.2.3 Search Service Parameters:

The Type-3 query defined in the WAIS-88 profile is being dropped in the WAIS-92 profile in favor of the Type-1 query which is now fully supported. The functionality that was previously defined by the Type-3 query is being re-implemented in the Type-1 query along with the definition of three new Bib-1 attributes; Any, Local Number and Doc ID.

5.2.4 Attribute Sets:

The WAIS-92 profile only requires the support of the three newly defined Bib-1 attributes, but support for other Bib-1 attributes can be added if needed. Support for the Title and Keywords attributes will have to be added to the WAIS-92 profile for conformance with the GILS profile.

5.2.5 Diagnostic Messages:

While the WAIS-92 profile is currently "silent" on the issue of diagnostics, it is expected to support the Bib-1 diagnostic set required by Z39.50, Version 2.

5.2.6 Record Syntax:

The WAIS-92 profile supports the Generic Record Syntax (GRS) but is "silent" on support for Unstructured Text (SUTRS) records and USMARC records. Moreover the WAIS-92 profile supports the Schema-1 schema defined for GRS records [9] which make it compliant with the GILS profile in that respect (which requires support for the Schema-1 schema). Support for SUTRS records and USMARC records [10] would need to be added to the WAIS-92 profile for full compliance with the GILS profile.

5.2.7 Preferred Presentation Format:

The preferred presentation format for SUTRS records is a Server implementation issue and is not part of the protocol.

5.2.8 Schema:

The WAIS-92 profile supports the Schema-1 schema for GRS records so is compliant with the GILS profile requirements in this respect [9].

5.2.9 Element Set Names:

The WAIS-92 profile supports both the "F" and the "B" element sets required by GILS (Z39.50, Version 2 requires the support of the "F" element set as a minimal requirement). The WAIS-92 profile supports the Title, Control Identifier, Dynamic Record Identifier, Originator and Date of Last Review elements in the "B" element set, and supports the Locator Record in the "F" element set as well as all the elements in the "B" element set (the "B" element set is usually a subset of the "F" element set). The mapping between GILS element names and WAIS-92 element names could be done as follows:

GILS Element	WAIS-92 Element
Title	Headline
Control Identifier	Record Identifier
Dynamic Record Identifier	URx
Originator	Name
Date of Last Review	Date
Locator Record	Object Element

However there is no support for the "G" element set in the WAIS-92 profile. As a result there is no support for the GILS Cross Reference and GILS Linkage

elements and this would need to be added for conformance with the GILS profile.

5.2.10 The Well-Known Search:

While there is no explicit support in the WAIS-92 profile for a "well-known" search, this could be implemented easily using the GILS profile attribute set Use Attribute: Doc Id; Structure Attribute: URL; and a term of zero length as described in the GILS profile.

5.3 Data Elements in the Locator Records:

The various data elements in the GILS profile can be supported by the GRS records using the Schema-1 schema, which is supported in the WAIS-92 profile [9].

6. WAIS-88 and GILS, Conclusions:

While the GILS profile could be implemented using the WAIS-88 profile, there would be a number of problems which would prevent the achievement of an ideal implementation. Notably the direct lack of support for Local Control Numbers, Control Identifiers and Availability Elements, more formalized support for hierarchical menu browsing (and "well-known" searches) and the lack of support for structured records such as USMARC records and GRS records (at a WAIS-88 profile level) would make such an implementation technically feasible, but not ideal.

7. WAIS-92 and GILS, Conclusions:

The GILS profile could be easily implemented under the WAIS-92 profile. However a number of areas still need to be addressed, these include: Availability Elements, more formalized support for hierarchical menu browsing (and "well-known" searches) and the "G" element set, support for SUTRS records and USMARC records. If these areas are addressed, then the WAIS-92 profile would be a suitable candidate for a GILS implementation.

8. WAIS-88 and WAIS-92 profile features which could be incorporated into GILS:

The following features of the WAIS-88 and WAIS-92 profile could be incorporated into GILS:

- Both the WAIS-88 and WAIS-92 profiles support the notion of "relevance ranking" of documents retrieved in a search. This feature allows the server to provide the Client (and the user) with an indication of the degree of "relevance" of each document in response to a search issued by the user.

- The GILS profile does not support the notion of "free-text" queries, only two Bib-1 attributes in the Type-1 query are supported which makes for very structured search. Support for a more unstructured query like that defined in the WAIS-92 profile could be considered. The reason for this suggestion is that there has been a lot of research done which indicates that casual end users, such as the general public, favor the use of unstructured queries rather than more structured queries, which are generally favored by professional searchers.

- Support for relevance feedback could be investigated further. This is one of the under-rated strengths of the WAIS system as a tool for casual end users searching. Again the reason for this suggestion is that there has been a lot of research done which indicates that the use of relevance feedback is an effective searching tool for casual end users.

- Both the WAIS-88 and WAIS-92 profiles support the notion of record "variants" (supported by Document Types in the WAIS-88 profile and by the "V" element set in the WAIS-92 profile). This feature allows the representation of records in a number of different formats and/or languages. Support for that feature would enable the support of languages other than English (such as Spanish for example) for the storage and display of records.

Notes:

1. Z39.50, Version 2, was approved in 1992. Since that time, work has continued to enhance the standard based on the needs of information providers. There is currently a draft Version 3 which is expected to be balloted through the National Information Standards Organization in 1994. The new version of the standard is referred to as Z39.50 -- 1994 and will describe both Version 2 and Version 3 of the standard. (Drafts of the new version of the standard can be retrieved from the Library of Congress's gopher. Connect to MARVEL.LOC.GOV and select #7. Services to Libraries and Publishers, and then select #8. Z39.50.)

2. The authoritative document describing the objectives, characteristics, etc. of the GILS is _Government Information Locator Service (GILS)_. Eliot Christian, United State Geological Survey, has been developing this document over the past few months. Various drafts have been available to the public for comment. The current draft dated January 22, 1994 is available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

3. The authoritative document describing the GILS profile at a technical level is _Using Z39.50 in an Application for the Government Information Locator Service: A Background Paper (January 24, 1994 Draft)_. This document has been under development over the past few months and has been recently released for public comment. The complete document is available on the Internet via anonymous FTP (File Transfer Protocol) from 128.230.33.81 as /USGS/gils_profile.txt (ASCII text format). After connecting to the FTP site, change directory to /USGS to retrieve the document.

4. This is very important point to bear in mind while reading the report. While a protocol may support a large range of features and capabilities, the user will not derive full benefits from them if they are not supported by both Clients and Servers. In that respect good Client and Server implementations are as important as the protocol they use to communicate with each other.

5. The Internet Engineering Task Force is a body involved in defining standards on the Internet. They are responsible for developing standards that support network application such as electronic mail, file transfer, and remote login, and the TCP/IP protocol suite that supports the Internet.

6. Universal Resource Locator and Universal Resource Name are a scheme currently being evolved by the IETF to allow the tracking and referencing of resources on the Internet (resources including documents, applications, services such as WAIS, Gopher, anonymous FTP, etc.). This scheme is under construction at this time and is still evolving.

7. The Internet Engineering Task Force is a body involved in defining standards on the Internet. They are responsible for such standards as Email, anonymous FTP, the TCP/IP protocol suite.

8. The Termination Services is not really a in Z39.50 service per se. It allows the Client or the Server to terminate the connection in an abrupt or graceful manner which is mapped to TCP ABORT and TCP CLOSE respectively (see Lynch, 1993).

9. GRS supports the use of numeric tags to identify the different fields in a record, however unless a Client "knows" the meaning of the numeric tags, it will be unable to assign a context to the different fields in that record and will merely know that the fields exist. This means that both Server and Client must use a previously defined GRS schema if numeric tags are to be used so that context may be assigned to the different fields. This is important because currently there is no provision for the negotiation of schemas between Clients and Servers in Z39.50. Where a pre-defined schema is not available GRS will support the use of text string tags (as a substitute for numeric tags), and these string tags could be used by the Client to assign context to the different fields in the record. Very recently a schema was defined for GRS records called Schema-1. The GILS profile requires support for this newly defined GRS schema which has also been adopted by the WAIS-92 profile. In that respect both profiles are compliant with each other.

10. USMARC is the implementation in the United States of ANSI Z39.2, Bibliographic information interchange. See American National Standards Institute (1985)

References:

Using Z39.50 in an Application for the Government Information Locator Service: A Background Paper (January 24, 1994 Draft). Available by anonymous FTP (File Transfer Protocol) via the Internet at 128.230.33.81 as /USGS/gils_profile.txt (ASCII text format).

American National Standards Institute. (1985). _American national standard Z39.2-1985: Bibliographic Information Interchange_. New York: American National Standards Institute.

Berners-Lee, Tim (1993, July) _Internet Draft, "Uniform Resource Locators"_.

Borenstein, N; Freed, N. (1993, September). _RFC 1521, "MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies"_.

Christian, Eliot. (1994, January 22). _Government information locator service (GILS)_. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

Franklin David, Brewster Kahle, Harry Morris, Jim Salem, Tracy Chen, Thinking Machine Corporation, Rod Wang, John Sui, Mark Grinbaum, Dow Jones & Company, Inc. (1990, April 23). _WAIS Interface Protocol, Prototype Functional Specification_. Available by anonymous FTP (File Transfer Protocol) via the Internet at wais.com as /pub/doc/protocol/protspec.txt.

Kahle, Brewster, Thinking Machine Corporation, (1990, May 5). _Document Identifiers or International Standard Book Numbers for the Electronic Age_. Available by anonymous FTP (File Transfer Protocol) via the Internet at wais.com as /pub/doc/protocol/doc_ids.txt.

Lynch, Clifford A. (1993, September 21). Using the Z39.50 Information Retrieval Protocol in the Internet Environment [Draft RFC for Z39.50 over TCP/IP].

McClure, Charles R., Ryan, Joe & Moen, William E. Moen. (1992). _Identifying and describing federal information inventory/locator systems: Design for networked-based locators_ 2 Vols. Bethesda, MD: National Audio Visual Center.

Morris, Harry, Thinking Machines Corporation (March 30th, 1990) _WAIS Protocol Users Manual_. Available by anonymous FTP (File Transfer Protocol) via the Internet at wais.com as /pub/doc/protocol/wisprot.txt.

National Information Standards Organization. (1992). _ANSI/NISO Z39.50-1992, Information retrieval application service definition and protocol specifications for open systems interconnection_. Gaithersburg, MD: National Information Standards Organization Press.

Weider, Chris and Deutsch, Peter. (1993, October) _Internet Draft, "Uniform Resource Names"_.

Appendix A:

IETF IIIR Working Group
INTERNET--DRAFT
Category: Informational
November 1993

M. St. Pierre, WAIS Inc
J. Fullton, CNIDR
K. Gamiel, CNIDR
J. Goldman, Thinking Machines Corp
B. Kahle, WAIS Inc
J. A. Kunze, UC Berkeley
H. Morris, WAIS Inc
F. Schiettecatte, FS Consulting

WAIS over Z39.50-1988

1. Status of this Memo

This memo provides information for the Internet community. This memo does not specify an IAB standard of any kind. Distribution of this memo is unlimited.

This document is an Internet Draft. Internet Drafts are working documents of the Internet Engineering Task Force (IETF), its Areas, and its Working Groups. Note that other groups may also distribute working documents as Internet Drafts.

Internet Drafts are draft documents valid for a maximum of six months. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a "working draft" or "work in progress."

Please check the I-D abstract listing contained in each Internet Draft directory to learn the current status of this or any other Internet Draft.

This Internet Draft expires May 1, 1994.

2. Introduction

The network publishing system, Wide Area Information Servers (WAIS), is designed to help users find information over a computer network. The principles guiding WAIS development are:

1. A wide-area networked-based information system for searching, browsing, and publishing.
2. Based on standards.
3. Easy to use.
4. Flexible and growth oriented.

>From this basis, a large group of developers, publishers, standards bodies, libraries, government agencies, schools, and users have been helping further the WAIS system.

The WAIS software architecture has four main components: the client, the server, the database, and the protocol. The WAIS client is a

user-interface program that sends requests for information to local or remote servers. Clients are available for most popular desktop environments. The WAIS server is a program that services client requests, and is available on a variety of UNIX platforms. The server generally runs on a machine containing one or more information sources, or WAIS databases. The protocol, Z39.50-1988, is used to connect WAIS clients and servers and is based on the 1988 Version of the NISO Z39.50 Information Retrieval Service and Protocol Standard. The goal of the WAIS network publishing system is to create an open architecture of information clients and servers by using a standard computer-to-computer protocol that enables clients to communicate with servers.

WAIS development began in October 1989 with the first Internet release occurring in April 1991. From the beginning, WAIS committed to use the Z39.50-1988 standard as the information retrieval protocol between WAIS clients and servers. The implementation is still in use today by existing WAIS clients and servers resulting in over 50,000 users of Z39.50-1988 on the Internet.

3. Purpose

The purpose of this memo is to initiate a discussion for a migration path of the WAIS technology from Z39.50-1988 Information Retrieval Service Definitions and Protocol Specification for Library Applications [1] to Z39.50-1992 [2] and then to Z39.50-1994 [3]. The purpose of this memo is not to provide a detailed implementation specification, but rather to describe the high-level design goals and functional assumptions made in the WAIS implementation of Z39.50-1988. WAIS use of Z39.50-1992 and Z39.50-1994 standards will be the subject of future RFCs.

4. Historical Design Goals of WAIS

As an aid to understanding the original WAIS implementation and its use of Z39.50-1988, the historical design goals of WAIS are presented in this section. Included with each goal is a brief description of the assumptions used to meet these design goals.

1. Provide users access to bibliographic and non-bibliographic information, including full-text and images.

Because Z39.50-1988 grew out of the bibliographic community, additional assumptions with the protocol were required to serve non-bibliographic information. They were also necessary to serve documents existing in multiple formats (e.g., rtf, postscript, gif, etc).

2. Keep the client/server interface simple and independent of changes in the functionality of the server.

To achieve this, the text string entered by the user was transmitted to the server without parsing the string into a Type-1 RPN (reverse-polish notation) query, as is common for bibliographic applications. Instead WAIS defined a new Type-3 query containing the text string. In this way, knowledge of the Z39.50 Attributes supported by the server was no longer required by the client or the user, as is true of many existing

Z39.50 implementations. In addition, the client software did not require modification to support the evolving functionality of the server.

3. Provide relevance feedback capability.

Relevance feedback is the ability to select a document, or portion of a document, and find a set of documents similar to the selection. WAIS included documents used in relevance feedback as part of the Type-3 query.

4. Permit the server to operate in a stateless manner.

A WAIS server was designed to be "stateless", meaning that search result sets were not stored by the server. In Z39.50 terms, the server exercised its right to unilaterally delete a result set as soon as it sent the search response. For this reason, the Present Facility of Z39.50 was not used, and retrievals were performed using the Search Facility. Relaxing this constraint in future implementations may prove the most prudent path.

5. Provide the ability for a client to retrieve documents in pieces.

Because retrieval of a portion of a document could be done several ways with Z39.50-1988, specific assumptions were made to implement this functionality. Accessing a portion of a document was required for both retrieval and for relevance feedback.

6. Run over TCP.

The Z39.50-1988 standard was designed to run in the application layer using the presentation services provided by the Open Systems Interconnection (OSI) Reference Model. Due to the popularity of TCP/IP and the Internet, WAIS was designed to run over TCP. Use of Z39.50 over TCP is described in [4].

5. WAIS Implementation of Z39.50-1988

By working with the Z39.50 Implementors Group (ZIG), the WAIS developers used a recommended subset of Z39.50-1988 and specific assumptions to fulfill its requirements. Over time, many of these requirements have then gone into the definition of subsequent versions of Z39.50. As new requirements become apparent, WAIS will document any additional assumptions and work with the ZIG in developing extensions.

WAIS supported the Init and Search Facilities of Z39.50-1988. Both search and retrieval were implemented using the Search Facility, as described in this section.

Search was initiated by the client with a Search Request APDU (Application Protocol Data Unit) using a Type-3 query. The query contained two main fields:

1. The "seed words", or text, typed by the user.
2. A list of document objects, where a document object is a full document, or portion thereof, to be used in relevance feedback. Each document object contains a document identifier (Doc-ID), type, chunk-code, and start and end locations. The Doc-ID and

type specify the location and format, respectively, of the document. The chuck-code determines the unit of measure for the start and end locations. Examples of chunk-codes used include byte, line, paragraph, and full document. If the chunk code is a full document, the start and end locations are ignored.

A Search Response APDU returned by the server contained a relevance ranked list of records, or WAIS Citations. A WAIS Citation refers to a document on the server. Each WAIS Citation contains the following fields:

1. Headline - a set of words that convey the main idea of the document.
2. Rank - the numerical score of the document based on its relevance to the query, normalized to a top score of 1000.
3. List of available formats - e.g. text, postscript, tiff, etc.
4. Doc-ID - the location of the document.
5. Length - the length of the document in bytes.

The number of WAIS Citations returned was limited by the preferred message size negotiated during the Init.

Retrieval of a document was initiated by the client with a Search Request APDU using a Type-1 query. The query contained up to four terms:

1. Term: Doc-ID
   Use Attribute: system-control-number — code = "um"
   Relation Attribute: equal — code = "re"
2. Term: the requested document format
   Use Attribute: data-type — code = "wt"
   Relation Attribute: equal — code = "re"
3. Term: the start location
   Use Attribute: paragraph, line, byte — code = "wp", "wl", "wb"
   Relation Attribute: greater-than-or-equal — code = "ro"
4. Term: the end location
   Use Attribute: paragraph, line, byte — code = "wp", "wl", "wb"
   Relation Attribute: less-than — code = "rl"

Because full-text and images were often larger in size than the receive buffer of the client, clients were designed to optionally retrieve documents in chunks, specifying the start and end positions of the chunk in the query. An example of a fully-specified retrieval query is:

```
query = ( ( use = "um", relation = "re", term = <Doc-ID> )
          AND
          ( use = "wt", relation = "re", term = postscript )
          AND
          ( use = "wb", relation = "ro", term = 0 )
          AND
          ( use = "wb", relation = "ro", term = 2000 )
        )
```

A retrieval response was issued by the server with a Search Response APDU. In this case a single record corresponding to the requested document, or portion thereof, was returned in the specified format.

6. Security Considerations

This RFC raises no security issues.

7. References

[1] National Information Standards Organization (NISO). American National Standard Z39.50, Information Retrieval Service Definition and Protocol Specifications for Library Applications, New Brunswick, NJ, Transaction Publishers; 1988.

[2] ANSI/NISO Z30.50-1992 (version 2) Information Retrieval Service and Protocol: American National Standard, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection, 1992.

[3] Z39.50 Version 3: Draft 8", October 1993. Maintenance Agency Reference: Z39.50MA-034

[4] Internet Draft, "Using the Z39.50 Information Retrieval Protocol in the Internet Environment", Clifford Lynch, November 1993.

8. Author's Address

Name: Margaret St. Pierre
Affiliation: WAIS Incorporated
Address: 1040 Noel Drive
Menlo Park, California 94025
Phone: (415) 327-WAIS
Fax: (415) 327-6513
EMail: saint@wais.com

Name: Jim Fullton
Affiliation: Clearinghouse for Networked Information Discovery & Retrieval
Address: 3021 Cornwallis Road
Research Triangle Park, North Carolina 27709-2889
Phone: (919)-248-9247
Fax: (919)-248-1101
EMail: jim.fullton@cnidr.org

Name: Kevin Gamiel
Affiliation: Clearinghouse for Networked Information Discovery & Retrieval
Address: 3021 Cornwallis Road
Research Triangle Park, North Carolina 27709-2889
Phone: (919)-248-9247
Fax: (919)-248-1101
EMail: kevin.gamiel@cnidr.org

Name: Jonathan Goldman
Affiliation: Thinking Machines Corporation
Address: 1010 El Camino Real, Suite 310
Menlo Park, California 94025
Phone: (415) 329-9300 x229
Fax: (415) 329-9329
EMail: jonathan@think.com

Name: Brewster Kahle
Affiliation: WAIS Incorporated
Address: 1040 Noel Drive
Menlo Park, California 94025
Phone: (415) 327-WAIS
Fax: (415) 327-6513
EMail: brewster@wais.com

Name: John A. Kunze
Affiliation: UC Berkeley
Address: 289 Evans Hall
Berkeley, California 94720
Phone: (510) 642-1530
Fax: (510) 643-5385
EMail: jak@violet.berkeley.edu

Name: Harry Morris
Affiliation: WAIS Incorporated
Address: 1040 Noel Drive
Menlo Park, California 94025
Phone: (415) 327-WAIS
Fax: (415) 327-6513
EMail: morris@wais.com

Name: Francois Schiettecatte
Affiliation: FS Consulting
Address: 235 Highland Avenue
Rochester, New York 14620
Phone: (716) 256-2850
EMail: francois@wais.com

Appendix B:

WAIS PROFILE OF Z39.50 Version 2
Version 1.2

FACILITIES AND SERVICE

For conformance with Z39.50-1994, the Initialization, Search, and Present Facilities are required.

ATTRIBUTE COMBINATIONS

Other Bib-1 attributes may be supported, but the following are required:

Use:	Any
	Local number
	Doc-Id
Relation:	Equal
	Relevance
Structure:	Free form text
	Document Text
	Local number
	URx

This profile assumes the following functions, to be supported by WAIS systems.

- o Free-Form (human-entered) Text.
- o Relevance Feedback by Document Text.
- o Relevance Feedback by Document Identifier.
- o Relavance Feedback by Record Identifier.
- o Search by Document Identifier.
- o Search by Record Identifier.

WAIS systems will use the following attribute combinations for these functions.

- o Free-Form (human-entered) Text.
  - Use: Any
  - Relation: Relevance
  - Structure: Free form text
- o Relevance Feedback by Document Text.
  - Use: Any
  - Relation: Relevance
  - Structure: Document Text
- o Relevance Feedback by Document Identifier.
  - Use: Any
  - Relation: Relevance
  - Structure: URx
- o Relevance Feedback by Record Identifier.
  - Use: Any
  - Relation: Relevance
  - Structure: Local number
- o Search by Document Identifier.
  - Use: DocId
  - Relation: Equal

Structure: URx
o Search by Record Identifier
Use: Local number
Relation: Equal
Structure: Local number

RECORD SYNTAX and SCHEMA DEFINITION

Support for Generic Record Syntax (GRS-1) is required.

WAIS records contain elements with numeric tags, except for "Object Element" which has a string tag. Tag-Type 1 refers to tags from Schema-1, and Tag-Type 2 are local tags.

Tag Type/Tag	Element	Repeatable?	Recommended Data Type
2/1	Headline	No	GeneralString
2/2	Name [optional]	Yes	GeneralString
1/16	Date [optional]	No	GeneralizedTime
1/?	Rank [optional]	No	Integer
1/13	Score [optional]	No	Integer
1/14	Record Identifier	No	GeneralString
1/12	URx[optional]	Yes	GeneralString
(see note 1)	Object Element		(see note 2)

Notes:
(1) Object Element may recur. Each occurrence will have a different string tag.
(2) The data type of Object Element might depend on the variant used. The datatype could be OCTET STRING, GeneralString, or EXTERNAL.

The Headline is set of words that conveys the main idea of the record.

The Name (optional element) is one or more individuals associated with the Document Object elements; it could, for example, be an author or an organization. As another example, there might be three Document Object elements: a fingerprint file, photo, and resume; the Name would be the individual that these Document Object elements describe.

The Date (optional element) is a date associated with the record.

The Score (optional element) representing the numerical score of the record based on its relevance to the query, from 1 to 1000, inclusive.

The Record Identifier is an identifier of that record, defined by, and unique within the target system (it may, but need not be, a URx).

The URx is a Universal Resource Locator or Universal Resource Name for the record.

Each Object Element contains object information of the record. It may be text, image, etc. Each instance may be available in one or more variants. Each instance has a different string tag. As in the above example, suppose there are three Document Object elements: a fingerprint file, a photo, and a resume; the string tags for these three elements respectively might be "fingerPrint", "photo", and "resume". It is recommended that the string tags for Document Object elements be descriptive of the information in the element, for discovery purposes (see description of element set "V" below).

ELEMENT SET SPECIFICATION

The following element set names should be supported. The latter three are supported with GRS.

- The primitive element set name "F" is required for conformance. The target will return all elements of the record as specified by the schema, where each Object Element is a default variant determined by the target. Large records may not be retrievable using this element set name.

- The primitive element set name "B" is used to obtain elements with numeric tags.

- The primitive element set name "V" is used to obtain elements with numeric tags, the string tag skeleton, and variant information, for each Document Object element: the string tag will be indicated for each Object Element, and for each, each variant for that element will be listed, as well as a variant identifier for each variant.

- The primitive element set name <variant identifier> (as returned by the target, as a result of a request which use the element set "V") may be used, to mean: return the "first portion" of the variant of the element associated with this variant id; "first portion" means from the beginning, as much as fit within the PDU.

- The primitive element set name <target position> (as returned by the target in a Generic record) may be used, to mean: return the "next portion" of the element and variant associated with this target position; "next portion" means beginning immediately after the end of the previous portion, as much as will fit within the PDU.

Interoperability and Conformance Issues in the Development and Implementation of the Government Information Locator Service (GILS)

Cecilia M. Preston and Clifford A. Lynch

May 22, 1994

INTRODUCTION

The Government Information Locator Service (GILS), as described by Eliot Christian [Christian, 1994], defines a initiative to implement a distributed system of autonomous, cooperating database servers attached to the Internet that provide locators for federal government information resources. Users of the GILS locate government information by retrieving records from these database servers; such searching is accomplished by client software that may run locally on workstations at GILS user sites or on host machines accessible to GILS users through the Internet. Government agencies participating in the GILS program will develop or acquire appropriate server software conforming to the GILS profile [McClure & Moen, 1994a] which will make their locator records accessible through the GILS. This server software will typically run on federal agency and other federal government computers[1].

The client software base used to access these GILS databases will be more heterogeneous in nature and diverse in origin. It will include client software designed specifically to provide access to the GILS; such client software, which might be developed either by the private sector and/or the federal government, is likely to be the most capable, incorporating the ability to navigate transparently among multiple GILS locator databases on behalf of the user, as well as perhaps the ability to connect users to at least some resources once located through a GILS locator. Because the GILS system is based on the American National Standards Institute (ANSI)/National Information Standards Organization (NISO) Z39.50 protocol for information retrieval [NISO Z39.50-1992], other client software already in place or under development for other purposes (for example, clients developed for access to the Wide Area Information Server (WAIS) system [Kahle et. al., 1992], or clients developed to communicate with bibliographic databases)[2] should be able to provide at least some access to GILS locator resources immediately even without explicit knowledge of the GILS profile. It is hoped that over time the capabilities of this already-existing software base will be upgraded to provide more extensive support of the GILS through explicit inclusion of support for features defined in the GILS profile.

As part of the GILS program, technical work was carried out by a group of experts led by William Moen and Charles McClure of Syracuse University to develop technical specifications for the GILS. This work resulted in a formal applications profile [McClure & Moen, 1994a], a standards document that is making its way through the National Institute of Standards and Technology's OSI Implementor's Workshop (OIW) under the auspices of the OIW Library Automation group and will likely ultimately become a Federal Information Processing Standard (FIPS). The applications profile is supplemented by several technical papers [McClure & Moen, 1993; McClure & Moen 1994b]. This effort focused on:

- The development of an architectural model for the GILS system and definition of participant roles and responsibilities for record creation and propagation within the

BEST COPY AVAILABLE

GILS. This was based on the functional specifications defined in [Christian 1994] and also drew from prior work on locator systems such as [McClure et al., 1992].

- The definition of data elements, interchange formats, and semantics for the locator records that form the contents of the GILS.

- The definition of the computer-to-computer communications protocols that are used to search and retrieve records from GILS servers. GILS uses a layered suite of protocols. At the lower layers, GILS uses the standard Transmission Control Protocol/Internet Protocol (TCP/IP) that is ubiquitous throughout the Internet; on top of TCP/IP the GILS employs Z39.50, the ANSI/NISO standard for computer-to-computer information retrieval[3]. Z39.50 is a very general purpose protocol, and Z39.50 Applications Profile (essentially, a specialization or restriction of Z39.50) is used to define the specific Z39.50 functions and parameters that are required to implement the GILS.

GILS clients and servers will be developed by many different organizations; a marketplace (in the broad sense of both public domain and commercial server and client software) that explicitly supports the GILS profile is expected to come into being as the GILS initiative moves forward within the federal government. In addition, a specific goal of GILS design is that GILS servers be usable, at least at a limited level, by a substantial base of already existing and deployed client software that supports the Z39.50 protocol; this will greatly enhance the availability of information in the GILS for the general public, particularly in the early stages of the initiative. Because of these objectives, the ability of a diverse base of clients and servers to work together, or interoperate, successfully is a central concern in the development of the GILS specifications and the implementation of the GILS. If agencies implementing GILS servers and users that wish to employ various GILS software clients to access them cannot have a reasonable expectation of such interoperability[4], it is likely that the GILS enterprise will fail. Further, if non-GILS Z39.50 clients cannot also interoperate successfully with GILS servers, the impact of the GILS effort in facilitating access to government information resources will be lessened.

This paper discusses various approaches that can be taken to increase the likelihood of successful interoperation among GILS servers and a variety of clients, including both clients designed to implement the GILS profile and other Z39.50 based clients. It includes both a discussion of the theoretical frameworks employed by standards development organizations to address interoperability considerations and also (and perhaps more importantly) approaches and experience by various implementor communities, such as the Z39.50 implementors, in the development of interoperable distributed systems and applications. The emphasis here, however, is not on the technologies and methodologies of interoperability or conformance testing, but rather what these approaches can contribute to the success of the GILS effort. In this connection, a number of additional means of promoting the development of interoperable systems, such as testbeds and reference implementations, are also discussed.

It is vital to recognize that GILS is a *system* specification, not just a specification for a Z39.50 applications profile. In this sense the use of the term "profile" may be somewhat confusing or misleading; GILS goes far beyond the usual sort of profile covered by an International Standardized Profile (ISP), for example (see [Ledrick & Spring, 1990] for a discussion of ISPs). In our view, this is a real strength of the GILS technical work: it is focused on providing a comprehensive, pragmatic blueprint for real interoperability that also takes account of the context of a large installed base of related systems. The GILS specification weaves together the use of several standards and specifies how they interrelate. The GILS specifications also address record content and its meaning. Because

of this broad scope, this paper includes a section on issues related to information semantics, interoperability, and quality as well as the coverage of system and protocol interoperability considerations.

CONFORMANCE TESTING AND INTEROPERABILITY TESTING

Basically there are two approaches to testing implementations to ensure that they work effectively together. One approach is *conformance testing*, in which a single implementation is compared to the standard to be sure that the implementation does what the standard specifies; the theory behind conformance testing is that if implementations all conform to the abstract standard they should interoperate with each other, although in practice this is not necessarily the case, as discussed below. The other approach is *interoperability testing*, in which two or more implementations are tested directly against each other, with the standard used as a primarily as a reference to adjudicate problems and incompatibilities, and secondarily as a guide to the functions to be tested and the general behavior to be expected. The objectives and expected results of these two types of testing are somewhat different. As Herbert Bertine and others define these differences: "Protocol specifications are used to develop products and services. Conformance testing verifies that these products and services comply with their specifications. Interoperability testing supplements conformance testing by verifying the end-to-end behavior of specified complex configurations". [Bertine et al.., 1990]

There is considerable debate about the value of conformance testing as a means of achieving interoperability. To some extent the disagreement has followed cultural lines, with the Internet community emphasizing a culture of running code and interoperability testing, and the traditional formal standards community (including the Open Systems Interconnection (OSI) developers operating within the framework of the national standards bodies such as ANSI and NISO in the United States and the International Organization for Standardization (ISO) internationally) and some academics advocating conformance testing as the primary approach, with interoperability testing as a relatively less important secondary activity.

> There is one school of thought which believes that in order to have true interoperability it is necessary to test each OSI implementation to ensure compliance to the appropriate standards and profiles...
>
> Many feel that conformance testing is not very effective. From a theoretical perspective, program verification is an unsolved problem, and it is simply beyond that state-of-the-art to produce programs that can guarantee that other programs are correct. There are plenty of instances of two implementations, each able to pass a conformance test, but not being able to interoperate with each other. Conformance testing therefore inspires little confidence in the user community. In contrast, conformance testing potentially provides benefits to an implementor during the early stages of development — as a simple check.
>
> From a practical perspective, conformance testing is probably the wrong approach. Consider: when an implementation is put under test, in effect it must interoperate with the test system. A conformance test is *nothing more* than an interoperability test with another implementation, the test system. The test system isn't placed in the user's environment — it isn't end-user equipment. The user doesn't buy test systems, the user buys systems to get work done. Interoperability testing against equipment the user will never buy — what's the point?...
>
> Interoperability testing is painful, but it is necessary. It is the *only* guarantee of working open systems. [Rose, 1990, p.588]

BEST COPY AVAILABLE

The recent Federal Internetworking Requirements Panel (FIRP) report [FIRP, 1994] has also endorsed the more pragmatically oriented approach of interoperability testing:

> However, the Panel is concerned that testing should be pragmatic (focused on demonstrating real interoperability), rather than theoretical (focused on conformance to specifications). Interoperability testing may consist of multivendor interoperability testing or interoperability testing against a reference implementation. [FIRP, 1994, section 4.5]

The FIRP report goes further, however, and begins to express a desire to attempt to link definitions of testing procedures, the performance of these procedures, and the interpretation, documentation and publication of their results to the procurement process for the federal government:

> Conformance testing can be a very valuable tool to the developer, but can be difficult and expensive to develop and execute definitively with rapidly evolving and integrated products. Instead, pragmatic tests that prove real world interoperability are required, to decrease the overall costs of testing and decrease the time to deliver tested products to market. On the other hand, once multivendor interoperability testing becomes the agreed criteria, agencies should insist on it, both through formal proof that products have indeed been tested against other products and reference implementations, as well as penalty clauses in system contracts for products that later prove not to be interoperable as advertised. NIST must help here by determining how to identify good, pragmatic interoperability tests and the appropriate procurement language to use them. Agency procurements should give preference to products which have demonstrated interoperability in accordance with the NIST program.
>
> For conformance or interoperability testing, test suites and approved means of testing must be available early, preferably at the same time as the standard or profile. Pragmatic testing criteria should be defined for all standards, including IPS and proprietary protocols. The cost of required testing must be proportionate to the value, and registers, of all tested products should be available in a publicly accessible on-line database. [FIRP, 1994, section 4.5]

The reader of the passage above should recognize that the FIRP report is specifying directions for new research and development efforts; the current state of the art in defining such tests is quite limited. Immediately after stating the objectives above, the FIRP report again returns to identify (and to a great extent to endorse) current pragmatic approaches such as testbeds (discussed in detail later in this paper):

> Recently, there have been a number of multi-vendor interoperability testing groups formed, often sponsored by an independent organization, and with significant end user participation (e.g., the FDDI interoperability test lab at the University of New Hampshire, the OSPF interoperability group). These groups focus on testing multiple vendors' implementations against each other, looking for bugs and areas of less than desirable robustness, etc. The Panel views these sorts of efforts as a step in the right direction by the vendor community. These types of practical testing efforts can improve the quality of real world implementations fielded by commercial vendors, especially when large end user organizations can also participate and bring test scenarios to the table. [FIRP, 1994, section 4.5]

This paper will examine the current state-of-the-art in both interoperability testing and conformance testing. Interoperability testing will be explored first; while it is a much broader perspective than pure conformance testing, current work in interoperability testing (and its limitations) provides a helpful context for understanding the even more serious limitations to the conformance testing approach.

Note that both conformance and interoperability testing can be performed by a number of different groups, including vendors and developers, system users (perhaps as part of a procurement process) or by neutral third parties that may offer some form of product certification or simply report test results back to the user and/or vendor communities, as suggested by the quotation from the FIRP report above.

Interoperability Testing and Its Limitations

While a precise definition of interoperability is somewhat elusive, functionally the meaning is clear: components of a system such as GILS communicate with one another effectively, correctly and provide the expected services to the user of a GILS client. In a very real sense, users don't care why components of a system like GILS fail to interoperate, or what component is at fault; while there can be many causes for failure, a successfully functioning operational system is clearly demonstrable to users. Further, users will view GILS as a totality; while there are a large number of standards and agreements involved in making the GILS work (each with its conformance and interoperability issues), users are only concerned that the entire constellation of standards, agreements and system components interoperate together effectively. While interoperability testing among implementors can address these concerns it is essential to recognize that they transcend individual standards, and hence the work that standards developers do on conformance and interoperability for individual specific standards cannot, by definition, address the full range of interoperability concerns that are central to the success of a system such as GILS.

Several other points about interoperability are of vital importance and need to be emphasized here; they all relate to the limitations of interoperability as a guarantee of developing a successful system. Interoperability is a necessary but not a sufficient criteria for successful development of distributed systems, and standards alone are not a sufficient basis for the development of a successful systems.

- Performance is a separate issue from interoperability. Systems may successfully interoperate but still suffer from devastating performance problems.

- Just because systems interoperate does not necessarily mean that they perform the functions that the user needs; it is entirely possible to identify and/or define standards and then develop a range of interoperable systems that implement these standards, only to discover that the system specification fails to successfully solve the problems or provide the functions that it was intended to provide due to errors in problem definition or solution specifications.

- Interoperability is concerned with communication between distributed computing systems; it does not speak to implementation issues such as user interface design or reliability of a given implementation. Interoperable implementations of a set of standards may still be rejected by their user community simply because they are badly implemented, hard to use or unreliable. Issues such as how well a given implementation provides help and diagnostic facilities to its users are not interoperability questions usually, but they are critical acceptance factors.

Testbeds as an Approach to Interoperability Testing

One approach that has been used successfully in distributed systems implementation is the testbed. Here a focused effort is made over a fairly short period of time to develop a number of implementations based on a set of standards that define a distributed system, and

to experiment with using these implementations to interoperate with each other. It is important to recognize that while one of the primary purposes of a testbed is to explore interoperability issues, a testbed typically takes on a broader role as a large scale experimental prototype for validating a system design.

Note that at least in our view testbeds imply active participation by software developers and vendors and standards developers, plus perhaps neutral organizers and facilitators and representatives of the user community; the focus is on system development and testing rather than simple validation and is perhaps most appropriate for products implementing relatively new standards. This is slightly different from the "multivendor interoperability testing groups" discussed by the FIRP report, which tend to be emphasize testing by third parties rather than the active participation of product and standards developers.

Testbeds were used in the early development of the TCP/IP protocol itself and as an aid to the development of mature, high quality implementation of this protocol; at that time they were called "connect-a-thons". The Z39.50 Interoperability Testbed (ZIT), sponsored by the Coalition for Networked Information (CNI) and involving about ten implementors of the Z39.50 protocol, played an important role in the development of interoperable Z39.50 implementations as the standard emerged in the marketplace in the early 1990s and also in identifying problems with the standard (in terms of ambiguous language in the standard, design problems in the standard, and missing functionality in the standard that was vital for real-world deployment of systems based on the standard).

Testbeds can be difficult to manage, however; they require substantial ongoing commitments of resources and energy on the part of the participants. Testbeds also call for a considerable level of trust and goodwill among the participants (particularly when the implementations involved include what will become commercial products, or are prototypes for potential commercial products from vendors that will later compete with one another in the marketplace, as opposed to experimental implementations from the research and education community) because of the need to share not only information about implementation but also because this is an environment that often reveals implementation problems not only to the implementor but to all participants in the testbed. Because of these issues, successful testbeds are typically organized under the auspices of some organization that is perceived as "neutral" by the participants (in the sense that the sponsoring organization does not favor one participant over another, and is not itself involved in the development of specific products for the marketplace, although it may well benefit from the successful creation of a range of high quality market offerings). A part of the sponsoring organization's role may be diplomatic, sorting out concerns and disagreements among participants. The sponsoring organization can also play a key role in publicizing standards and showcasing interoperable implementations for a possibly skeptical or poorly informed user community; because of its "neutral" position the organizer may be more credible to the user community than the participating vendors would be alone.

Because the emphasis is on implementations, testbeds lead to a "whole system" approach to testing rather than one focused on individual standards conformance or interoperability and can be very useful not only in dealing with problems directly related to a given standard but in identifying problems that arise from the interaction between different standards or at the boundaries between standards and implementor agreements often needed to produce real-world interoperating systems. They are also helpful in identifying potential performance problems, functional errors in problem specifications and protocol design, and in providing early warning of poor quality implementations. Testbeds are also valuable in providing implementors with insight into features of standards that may be problematic to implement and that can lead to unacceptable implementations, and can provide vital feedback to the standards development process, particularly if some of the standards developers are

involved in the testbed process. A successful testbed becomes an incubator for the development of a shared base of engineering know-how for standards implementation.

Implementations that have been proven in testbed activities tend to become robust quickly; they include provisions to detect and recover from erroneous information sent from other implementations and to provide extensive diagnostic tools (such as logging facilities) for debugging in these situations. One of the problems with protocol standards in general is that they usually don't address what to do in the case of errors such as incorrectly structured or sequenced protocol data units received from the remote system; immature or poorly designed implementations often typically *assume* that peer systems on the network are sending correctly coded and sequenced protocol data elements and do not include "suspicious" code to check for and recover from such errors. The system that freezes or crashes, is typically a result of a simplistic implementation encountering other implementations behaving incorrectly. This behavior is not accepted in high quality production implementations; graceful recovery from incorrect information sent from another system is vital in quality implementations (though it is not a standards issue). In a testbed environment implementations are typically exposed to a broad range of pathological behavior from other testbed participants, and thus are able to quickly achieve a high degree of robustness and maturity, which is difficult to obtain through other methods.

Testbeds can also yield dividends that will facilitate the overall development of a marketplace of products implementing a standard. Implementation knowledge gained by testbed participants can be used by other implementors later to speed up development and avoid errors, if this information is appropriately captured and shared. Note that information sharing of this type can be a major policy issue in organizing and managing a testbed project, and particularly in the context of adding new members once a testbed project is underway, or relating testbed activities to those of the broader implementor community outside of the testbed group, since commercial participants who are making the investment in testbed participation may regard the implementation engineering knowledge gained as a valuable asset which they are unwilling to share freely, especially with competitors that did not choose to invest resources in testbed participation. Testbed "pioneers" may not be satisfied with simply enjoying the early implementation lead that participation in a testbed can provide, and may wish to maintain that leadership position as long as possible.

Another common role of testbeds is as public demonstrations of the viability of a suite of standards that define a distributed system; they can be central not only in convincing a skeptical user community that distributed information systems can work, as discussed above, but also in helping the user community to understand the implications and operational limitations of such systems. For example, in the Z39.50 Interoperability Testbed project demonstrations provided the library community with the first real understanding of the implications of decoupling user interfaces (clients) from information servers and helped to advance consideration of the various policy, planning, and user support issues that such a decoupling raised for the library community.

It should be emphasized that while there is a good deal of consensus on the value of interoperability testing, either on a case by case basis or in broader contexts such as testbeds, there is very little methodology for how such interoperability testing should be accomplished. The FIRP report, for example, exhorts the National Institute for Standards and Technology (NIST) to devote efforts to defining such methodologies; this is a research problem, although the FIRP report does not explicitly identify it as such. Essentially, the idea behind interoperability testing is simply to exercise the software and systems involved as extensively and as stressfully as possible. In some cases emphasis has been deliberately placed upon stress conditions— for example, trying to send sequences to other systems that are legal within the protocol but represent boundary cases that might cause the other system

to crash, with prizes to the last system left standing. In other cases, such as the Z39.50 testbed, emphasis was also placed an just understanding the behavior of interoperating systems, and the limits of effective interoperability at the applications level, such as the various ways in which different systems might interpret a given query. Ultimately, successful interoperable testing relies upon sufficient time commitments by energetic testers — notably those very familiar with the standards in question, with specific implementations of the standards and also by those who represent a user perspective. This last group is difficult to enroll in a testbed project, since they don't have an obvious direct economic interest as the participating vendors do, though they are certainly members of a community with a strong interest in the successful product of the testbed project; these "end-user" representatives may require funding or other support to participate. But for applications-level standards such as Z39.50 their participation has been essential in successful testbed activities, and we believe that it will be important to engage them in any testbed that supports the GILS initiative.

CONFORMANCE TESTING: LIMITATIONS AND PROBLEMS

Interoperability testing is an art. It is a somewhat imprecise process that achieves its results through the active engagement of a group of implementors and other system testers sharing a common set of goals.

In contrast, there has long been a desire to develop a science (or at least an engineering discipline) of standards conformance testing. The analogy to software development here is useful. Computer scientists have performed research in proofs of program correctness as a rigorous science for decades (with very limited practical results); software developers meanwhile have developed a great deal of largely anecdotal and heuristic knowledge about how to test large-scale software programs and now normally include extensive testing and quality assurance programs (much more akin to interoperability testing) as part of their development cycles.

Rather more progress has been made in standards conformance testing than in the much more general (and hence complex) problem of proof of program correctness. There are still a number of major problems. For example, it is very difficult to rigorously test that an implementation conforms to a standard that it itself not rigorously specified. Thus it is fairly tractable to test areas of conformance such as correctness of state transitions where the desired behavior can be modeled in a reasonably clear, unambiguous and formal fashion, or to determine whether protocol data units generated by an implementation of a protocol are syntactically well-formed. Determining whether an implementation conforms to a specification that is only rather loosely defined in prose (as is typical of protocol semantics) is a much more intractable problem. In many key cases, it turns out that the standards are silent on the specific interpretations; for example, in Z39.50 Version 2 [Z39.50-1992] incompatibilities between implementations have arisen because of disagreements about the need for case sensitivity in the interpretation of database names and because of conflicting assumptions about the semantics of omitted or repeated attributes in search queries.

This difficulty becomes more acute as one moves up the layers of a hierarchical protocol suite. At the bottom levels of the protocol hierarchy (e.g. the data link layer) one is dealing with the rather mechanistic movement and processing of bits and bytes and models such as finite state automata can very precisely define the protocol's operation. For applications layer protocols, while some parts of the standard may lend themselves to such a mechanistic definition (and thus very clear conformance testing, and also the development of test suites that examine a series of key behaviors) such as the algorithms in the Z39.50 protocol that determine how a server blocks records into protocol data units in a PRESENT

RESPONSE, other parts of the standard address semantics of search processing and seem extremely resistant to abstract modeling. An excellent example of these problems is the ongoing disagreements within the Z39.50 implementor community about the extent to which it is appropriate for a server to apply liberal interpretations to attributes in queries when it does not support the precise attribute combination specified by the client.

Another limitation of conformance testing has to do with the generality of the protocols being tested. Particularly at the OSI model applications layer, protocols are typically very complex and general, offering many options and choices. There is often no reason to believe that even two completely correct and conformant implementations of an applications layer protocol will interoperate in any useful way. Thus conformance testing at this protocol level may be of extremely limited value. As is the case with GILS, a protocol standard (such as Z39.50) may be further constrained by an applications profile, and conformance testing against the combination of a protocol standard and an applications profile may be more useful. In the OSI world, an additional specification called a Protocol Implementation Conformance Statement (PICS) is sometimes used both to provide further constraints that define a particular class of implementations of a protocol and to document the ways in which each implementation varies from the combination of protocol standard and applications profile.

The International Organization for Standardization (ISO), which has invested a considerable effort in conformance testing methodologies and approaches, is clear about the limitations of not only the current state of the art but also the objectives of conformance testing as they view it. As stated in ISO/IEC 9646, which sets out the overall framework for specifying conformance test suites for OSI protocols and for defining the procedures to be followed during testing:

> Conformance testing involves testing both the capabilities and behaviour of an implementation, and checking what is observed against both the conformance requirements in the relevant International Standards or CCITT Recommendations and what the implementor states the implementation's capabilities are.
>
> Conformance testing does not include assessment of the performance nor the robustness or reliability of an implementation. It cannot give judgments on the physical realization of the abstract service primitives, how a system is implemented, how it provides any requested service, nor the environment of the protocol implementation. It cannot, except in an indirect way, prove anything about the logical design of the protocol itself.
>
> The purpose of conformance testing is to increase the probability that different OSI implementations are able to interwork. However it should be borne in mind that the complexity of most protocols makes exhaustive testing impractical on both technical and economic grounds. Also, testing cannot guarantee conformance to a specification since it detects errors rather than their absence. Thus conformance to a test suite alone cannot guarantee interworking. What is does do is give confidence that an implementation has the required capabilities and that its behaviour conforms consistently in representative instances of communication. [ISO/IEC 9646-1 p. v]

There is a substantial research literature on conformance testing (see, for example, [Bertine et al., 1990; Bush et al., 1990; EC 1991; Pink, 1990; Probert & Desjardins, 1990; Vermur & Blik 1993]) as well as the ISO/IEC 9646 document which defines the overall framework and specific standards documents addressing conformance tests for individual standards in the OSI context. ISO/IEC 9646 also defines a taxonomy of conformance requirements (static and dynamic) and of types of tests.

We will not attempt to summarize this material here; the interested reader is referred to ISO/IEC 9646.

There is a great temptation to pursue the definition of conformance testing for GILS. The presence of existing literature, standards documents, and taxonomies gives the illusion that one is doing well-defined, well-understood, rigorous engineering and making progress within a generally accepted context. However, as the discussion thus far emphasizes, the payoff from extensive investments in conformance testing may be quite limited. It certainly does not begin to solve the critical problem of insuring interoperability among GILS components and making the GILS a success.

REFERENCE IMPLEMENTATIONS AND TESTBEDS

Reference implementations can play a very important role in promoting the development of a critical mass of interoperable implementations of a standard or suite of standards. Some reference implementations have been explicitly developed, in the sense that an agency concerned with the deployment of a standard or standards suite has funded an implementation that was widely distributed. Arguably this was the case with TCP/IP, to the extent that an implementation was developed and widely distributed with Berkeley UNIX through funding from the Department of Defense Advanced Research Projects Agency (ARPA) and this TCP/IP implementation became a de facto reference implementation. In some cases, the reference implementation also serves as a code base from which other (commercial) implementations are developed.

In other cases, de facto reference implementations have evolved because a number of early implementors (all of who successfully interoperate with each other) have made servers available for new implementors to test against. Sometimes these de facto reference implementations will emerge almost by acclamation; they will be among the more robust implementations that may have been part of a testbed early in a protocol's development, or perhaps the most accessible implementations (in that they are publicly available for testing). The extent to which the organization providing such access to its implementation is prepared to offer debugging assistance to other implementors also often plays a role in establishing its implementation as a reference for the implementor community. In some cases a given organization's implementation may be so important to the marketplace that it becomes one of the de facto reference implementations because that organization can mediate disputes about conformance and interoperability by virtue of its marketplace position.

Note that source code or even object code (binaries) need not be made available in order for an implementation to serve in the de facto reference implementation role for complex application level standards like Z39.50 or GILS; the key point is that a working service is available on the network for testing new implementations against. This has certainly been proven to be the case with Z39.50 (particularly in bibliographic applications) where there are a number of publicly available servers accessible across the Internet for interoperability testing, such as OCLC, RLIN, DRA, the University of California and Pennsylvania State University. A developer that has tested interoperability against all of these systems successfully will not have perfect code, but can be assured of a reasonable stable and correct implementation of the Z39.50 protocol.

One of the major problems with getting successful, interoperable implementations of a new standard started is the creation of a group of de facto reference implementations, at least for testing purposes. This is essential if interoperability, rather than conformance testing is to be the primary approach. The formation of an testbed early in the adoption and implementation of a protocol can play an essential role in creating the initial core of de facto reference implementations to test against, and in helping other vendors who are considering

investments in developing products that implement the new protocol to commit resources to product development. It is important to ensure that access to the de facto reference implementations continues to be made available to new developers entering the marketplace even after the first wave of products have appeared, though this usually will occur naturally since easy accessibility is usually part of the characteristics that *define* the de facto reference implementations.

An advantage of using a testbed populated by a mix of experimental implementations from the research and education community and product prototypes from the commercial sector rather than commissioning a single reference implementation is that it can stimulate the development of industry-wide expertise without the problems of having a funding agency pick a single winner (thus giving the selected implementor of the reference implementation a privileged position for commercial competition) or allowing a single organization to dominate developments thorough its ability to interpret or even amend a standard. Without selecting a 'winner" (a designated reference implementation) funding from a sponsoring agency can still be useful in moving the testbed project forward and thus promoting implementation of the standard, and the creation of a marketplace of products that implement it. For example, funding might be provided to support testbed activities, the documentation and dissemination of knowledge gained about the protocols involved and the implementation issues involved in developing software that uses them, and perhaps even to ensure that one or more experimental implementations are available early and contain sufficient instrumentation to facilitate protocol analysis and testing. Funding can also be used to ensure that representatives of the user community participate in the testbed.

History suggests that if a single reference implementation must be funded it is vital to differentiate it from products that may later enter the marketplace. One approach has been to have the reference implementation done by an organization that will not compete in a commercial marketplace (such as a university or other nonprofit). The resulting implementation is then placed in the public domain, or licensed at low cost to all who are interested; in some cases it becomes a beginning code base that commercial implementors can build upon, with all of the interoperability advantages that a common initial code base offers[5]. The approach of commissioning a non-commercial reference implementation has been used with mixed success in a some cases such as X.500 directory services; a university or other research organization has been commissioned to develop the reference implementation and has used development tools and approaches that have emphasized flexibility and speed of implementation, and evaluation instrumentation, rather than the robustness and high performance that characterizes successful commercial implementations. This helps to differentiate the reference implementation from other commercial implementations, but it does have its dangers. The reference implementation may turn out to be widely used, simply because it is free, or be the first implementation that the user community gains much experience with. If this implementation is slow or unreliable it may lead to a perception by the user community that the *standard*, rather than the implementation, is irretrievably flawed and may thus play a role in rejection of the standard by the user community.

Well-known, readily accessible reference implementations on the network also serve an important function for customers. Rather than relying on vendor claims about interoperability or having to make complex, costly arrangements for trial implementations and acceptance testing, a potential customer can quickly gain a reasonable degree of confidence about the ability of a vendor system to interoperate simply by asking the vendor to demonstrate interoperation with one or more of the well-known reference implementations to the customer — this can even be done at a trade show if the vendor booths contain systems attached to the network.

The FIRP report endorses the idea of reference implementations and recognizes their importance (though it does not explicitly recognize the developing role of publicly accessible reference implementations as services on the network):

> Interoperable implementations of the standards and profiles must also be available, preferably with at least one reference implementation in the public domain. Standardization on technology which has yet to reach implementation and limited deployment stages has generally been less than successful. [FIRP, 1994, section 4.4]

Several comments should be made on the FIRP recommendations, however, in the context of GILS. First, GILS is a *profile* based on a set of well-established technology and standards, not an entirely new standard; Z39.50, for example, is already widely deployed and well-accepted. Thus there is already a sizable base of existing implementations, both commercial and public domain, that will interoperate with GILS implementations to at least a limited extent. Second, at least for network based information access applications, it may be that the FIRP report goes farther than necessary in calling for public domain implementations (although these do exist for Z39.50); perhaps more important, at least in our view, is that servers be publicly accessible on the network for interoperability testing, including testing by potential purchasers of commercial products.

CERTIFICATION

Certification — either of conformance or of interoperability — is an extremely attractive idea in the abstract; suppliers of servers and/or clients would receive some kind of certification from a testing agency, and purchasers of products that carry this certification could be assured that the products would interoperate with other products. Certainly, the FIRP report looks forward to a time when certification of products could be available and play a role in the federal procurement process.

This approach carries a number of political (and, very likely, legal) complexities. Who should perform such certification? How should the certification activities be funded? How are disputes adjudicated? What methodology and tests would be used to verify conformance or interoperability? In the case of interoperability, what other products should a given product be tested against?

Beyond the political and legal problems are very real technical ones. Certification is a much more comfortable fit with the rather mechanistic processes of conformance testing: for example, a certification that an implementation successfully processes a test suite of protocol data elements. This does not address interoperability issues (and indeed, it is not clear how one would certify interoperability, except against a limited number of specific systems at a specific point in time). It also does not provide a way to address semantic issues that are very important to the user community.

Finally, it should be noted that certification is not, in our view, particularly useful in moving a suite of standards and implementations towards maturity or in validating a new profile or suite of standards; it often fails to directly engage the key communities of implementors, standards developers and end users.

READY AVAILABILITY OF STANDARDS DOCUMENTS

The effect of having standards documents (both in draft form and in final versions) readily accessible in electronic form on the Internet for public use and review has proven to be of

substantial importance in furthering development of a base of interoperable implementations. The difficulty and high cost of identifying and acquiring relevant OSI standards for implementation efforts (or, indeed, for teaching or research purposes) has proven, over time, to be a substantial barrier to the marketplace adoption of these standards. Further, the difficulties in obtaining OSI standards have limited their review by interested communities that might have greatly improved both the quality and the comprehensibility of these standards. This is especially important for applications level standards due to the very broad community that is interested in them.

Internet standards (the so called Requests for Comments, or RFCs) , in contrast, are publicly available at no cost through the Internet, both as drafts during their development and later in final form. A typical Internet standard receives a very wide review from the potential user and implementor communities as part of its development. Internet standards also form a vital part of the base of material used in teaching and research. Relevant RFCs are available instantly to implementors who need to refer to them; thus they are used heavily to resolve questions during design and implementation.

We would argue, with Carl Malamud [Malamud, 1992] that the effects of easy public availability of both draft and final standards in electronic form have been underestimated as a factor both in the quality of standards and in acceptance and adoption of these standards by the implementor and user communities.

The Draft Report of the Federal Internetworking Requirements Panel has endorsed the Internet RFC model as the appropriate approach for federal networking standards:

> The Panel also believes that all standards and profiles used in federal networking need to be widely available in electronic and paper form at low or no cost. Consistent with the policy espoused in OMB Circular A-130, these fees should cover the cost of dissemination of the standard, not the cost of its development. [FIRP, 1994, section 4.4]

The GILS effort has followed the Internet RFC model in making its documents widely available for review and reference through the Internet. Indeed, the initiative has also reached out to a very broad community through the use of List Servers (LISTSERVs) and other electronic mail reflectors to distribute announcements about the availability of draft documents at various points, and by providing versions of these draft documents in a wide range of popular formats. Strategically, this has been an important decision, and one that can be expected to contribute to the development of a large base of interoperable implementations.

CONFORMANCE, INTEROPERABILITY AND DATABASE SEMANTICS

One of the difficulties of GILS from an interoperability testing point of view is that the profile and related documents specify not only computer-to-computer protocols but also discuss the content of locator databases. A well-constructed GILS client should understand, for example, how to interpret browsing menus and cross reference records and display them usefully to users; such a client should also understand the unique IDs present in GILS records and use them to eliminate duplicate records from result sets obtained by searching across multiple GILS servers before displaying these results to a user. These are not protocol functions, but rather capabilities that are available to the client because it understands the semantics of GILS records. When one speaks of conformance and interoperability in this context, one is speaking about client function and not about protocols. There is very little precedent for discussing system behavior in this context.

Yet a client system can interoperate (successfully interchange search requests and receive data) with a GILS locator database to the extent of searching the database and presenting records to the user without supporting any of these features, and without understanding the semantics of GILS records. Indeed, to facilitate limited interoperability with the installed base of bibliographic and WAIS clients GILS servers are expected to support forms of record export that to some extent conceal the full semantics of locator database records. This might be the case in communicating with an existing Z39.50 client designed to support bibliographic searching, for example. Here critical issues are the mapping of the GILS data elements into MARC records which the server will then present to the bibliographic client, and the quality of the resultant MARC records. In this context it makes little sense to discuss client conformance and interoperability because the clients were never designed to interoperate with GILS servers in the first place; rather the interoperability issues become those of having GILS servers successfully emulate the kinds of servers that these clients were designed to interoperate with, including the ability of GILS servers to perform semantic mappings of data elements from locator database records into record interchange formats that are know to this base of clients. The GILS servers must in fact conform to other standards (such as MARC) and interoperate with systems that implement these other standards. One open question is whether the MARC records exported by GILS servers will meet the minimum standards for completeness and quality that bibliographic clients may establish; this is particularly troublesome because there are no explicit standards for such MARC records broadly accepted within the bibliographic community.

The definition and description of these various levels and forms of interoperability are likely to be quite confusing to potential customers for GILS clients.

CONCLUSIONS AND RECOMMENDATIONS

Interoperability testing, rather than conformance testing, should be the keystone of any program to further the development of an interoperable base of GILS clients and servers. GILS, as an Internet application, has already positioned itself within this tradition, and has already made good use of practices within this tradition such as the public distribution of draft standards documents for wide review and comments. We do not recommend any substantial investment of effort in the development of conformance tests. Some highly non-comprehensive conformance test streams were developed within the Z39.50 implementor community as a debugging (rather than conformance testing) tool; these will be useful to GILS implementors, and it might be worth the rather minimal investment of effort to develop some collections of test protocol data units (PDUs) specifically as a debugging tool for GILS implementors.

Certification is not a feasible approach, either politically or technically, at this stage in the development of the GILS. Leaving aside the basic problems of certification, we would argue that the priority at present is to create a base of conformant implementations and of implementation expertise and to validate the GILS profile through implementation (make changes to the profile as required based on implementation experience). This is best achieved by actively engaging implementors, users and standards developers rather than by assigning responsibility for testing or certification to some third party.

It is essential to recognize that the success of the GILS initiative depends on more than simply interoperable software. The GILS system architecture and profile documents contain rules and guidelines for the construction of content for the locator databases. These need to be validated as well; a key issue will be the development of knowledge about how to construct locator databases, and the resolution of questions about how to propagate

records from one database to another, the appropriate inclusion of cross-reference records, and the granularity of information resources identified by locator records.

Interoperability testbeds, in part as a way of developing a de facto core of well-known reference implementations and in part as a way of simply moving early implementations to a more mature state and ensuring their mutual interoperability, deserves careful consideration. A testbed can also be justified as a means of validating the system architecture and the profile, and as a means of gaining experience with content-related questions. Testbeds have been an effective approach in other, similar enterprises, including the development of Z39.50 clients and servers for bibliographic applications. GILS has the advantage that many of the Z39.50 developers that would either produce GILS profile specific clients and servers or who would want to upgrade their existing clients to work well with GILS servers are already familiar with the testbed model and indeed may have had prior experience participating in testbeds. If the testbed approach is followed, one important decision would be the selection of an appropriate neutral party to organize the testbed. While we do not have a specific recommendation for a sponsoring organization for the testbed, we feel that it because so many of the testbed issues revolve around content and its representation that the organizer should not be a federal agency that has a vested interest in the way they have structured the contents of their locator database.

Another unique issue for the GILS project is that two closely linked testbed projects may be needed; one for implementors of the GILS profile proper, and a second to explore interoperability issues between GILS profile servers and the existing installed Z39.50 client base. Building these two testbeds and carefully articulating and demonstrating the differences between them might also help to address confusion that will occur in the user community about interoperability expectations for GILS profile and non GILS profile clients communicating with GILS servers.

Particular attention will need to be given to exploring interoperability issues (and related quality issues) in data content and data element mappings to record types such as MARC, as well as to defining and explaining the various forms of partial interoperability that may be achieved between GILS servers and clients developed by other communities of Z39.50 implementors, particularly the WAIS community and the bibliographic community. The bibliographic Z39.50 implementor and user communities are particularly important because they include the federal depository libraries and the broader community of government documents librarians, who are already familiar with bibliographic systems (including those which are used for organizing government documents) and who will also be an essential user community for the GILS, both as direct users and as intermediaries that assist the broader public in using the GILS. This is poorly explored and complex territory and to some extent falls outside of the usual framework of conformance and interoperability testing for protocols, yet it is critical to the success of the GILS initiative. We strongly recommend that if a testbed is pursued, these communities be brought into the testbed as representing one of the key user perspectives.

NOTES

[1] It should be noted that while the GILS documents only directly address the development of locator databases by federal government agencies, there is nothing in the architecture or design of the system that precludes the expansion of the GILS system to also incorporate locator databases providing access to information resources outside of the federal government or locator databases developed outside the federal government but providing access to a mixture of federal government and other information resources (for example, these locator databases might be developed by libraries, by commercial information providers, by state or local governments within the United States, or, indeed, even by other national governments or international organizations). In addition, the GILS architectural model could also be replicated by other organizations without any content linkage to the US federal government GILS system.

[2] For a sense of the scope and diversity of the existing Z39.50 client base, see [Moen 1994].

[3] Actually, because the use of Z39.50 is not defined in a TCP/IP environment by the ANSI/NISO Z39.50 standard, GILS also specifies the mapping of Z39.50 on top of TCP/IP [Lynch, 1994].

[4] While outside the scope of the GILS program, it seems both desirable and probable that clients designed specifically to support the GILS profile will also be designed to interoperate with other Z39.50 servers, such as WAIS servers or bibliographic servers.

[5] One of the issues with a common public domain or other generally available code base is that problems as a standard evolves, and the need for mechanisms to manage this code base. Various approaches have been taken to address this; for example, the code base of Berkeley UNIX was re-released for each new version of the de facto standard, which meant that commercial developers had to repeatedly re-integrate the common source code base. The X Consortium also uses a common code base, and invites implementors to contribute code back to the common code base that becomes part of new releases.

REFERENCES

Bertine, Herbert, Elsner, Wolfgang B., Verma, Pramode K., and Tevani, Kamlish T. "Overview of Protocol Testing Programs Methodologies and Standards" *AT&T Technical Journal.* (January/February 1990): 7-16.

Bush, Matthew, Rasmussen, Kris and Wong, Tai. "Conformance Testing Methodologies for OSI Protocols: *AT&T Technical Journal.* (January/February 1990): 84-100.

Christian, Eliot. *Government Information Locator Service (GILS); Report to the Information Infrastructure Task Force.* (May 2, 1994). Available by anonymous FTP from 130.11.48.107 as /pub/gils.doc (in Microsoft Word format) or /pub.gils.txt (ASCII text).

Conformance Testing and Certification in Information Technology and Telecommunications. Proceedings of the European Conference, Brussels, Belgium 13-15 June 1990. Amsterdam, Netherlands; IOS, 1991. (cited as [EC 1991])

Federal Internetworking Requirements Panel. *Draft Report of the Federal Internetworking Requirements Panel.* Prepared for the National Institute of Standards and Technology, January 14, 1994.

ISO/IEC 9646-1. International Standards Organization. *Information Technology — Open Systems Interconnection — Conformance testing methodology and framework. Part 1: General Concepts.* 1991.

ISO/IEC 9646-2. International Standards Organization. *Information Technology — Open Systems Interconnection — Conformance testing methodology and framework. Part 2: Abstract test suite specification.* 1991.

ISO/IEC 9646-3. International Standards Organization. *Information Technology — Open Systems Interconnection — Conformance testing methodology and framework. Part 3: The Tree and Tabular Combined Notation (TTCN).* 1992.

ISO/IEC 9646-4. International Standards Organization. *Information Technology — Open Systems Interconnection — Conformance testing methodology and framework. Part 4: Test realization.* 1991.

ISO/IEC 9646-5. International Standards Organization. *Information Technology — Open Systems Interconnection — Conformance testing methodology and framework. Part 5: Requirement on test laboratories and clients for the conformance assessment process.* 1991.

Kahle, Brewster, et al.., "Interfaces for Distributed Systems of Information Servers" *Networking, Telecommunications, and the Networked Information Revolution.* Proceedings of the ASIS mid-year meeting, May 28-30, 1992, Albuquerque, NM. Silver Springs, MD; ASIS, 1992. 124-148.

Ledrick, Diane P. & Spring, Michael B. "International Standardized Profiles". *Computer Standards & Interfaces* 11(1990): 95-103.

Lynch, Clifford A. *Using the Z39.50 Information Retrieval Protocol in the Internet Environment.* Draft Internet RFC.

Malamud, Carl. *Stacks: interoperability in today's computer networks.* Englewood Cliffs, NJ; Prentice Hall, 1992.

McClure, Charles R. and Moen, William E. *Application Profile for the Government Information Locator Service (GILS)* (May 7, 1994). Available by FTP from ericir.syr.edu. (Cited as [McClure & Moen, 1994a])

McClure, Charles R., and Moen, William E. *Using Z39.50 In An Application For The Government Information Locator Service (GILS): A Background Paper* (May 7, 1994) Available by FTP from ericir.syr.edu. (Cited as [McClure & Moen, 1994b])

McClure, Charles R. and Moen William E. *Expanding Research and Development of the ANSI/NISO Z39.50 Search and Retrieval Standard.* (October 16, 1993). Available by FTP from ericir.syr.edu.

McClure, Charles R., Ryan, Joe, and Moen, William E. *Identifying and Describing Federal Information Inventory/Locators Systems: Design for networked-based Locators.* 2 vols. Bethesda, MD; National Audio Visual Center, 1992. Available from ERIC document no. ED349031.

Moen, William E. *The Z39.50 Protocol: Information Retrieval in the Information Infrastructure.* (June 1994). Available from NISO.

NISO Z39.50-1992. National Information Standards Organization. *ANSI/NISO Z39.50-1992, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection.* Gaithersburg, MD; National Information Standards Organization Press.

Pink, Jane. "Conformance Testing". *An Analysis of the Information Technology Standardization Process.* John L. Berg and Harold Schumny (eds.) Elsevier Science Publishers (North-Holland), 1990. 111-116.

Probert, R.L. and Desjardins, M.M. "Improving Quality and Interoperability of Protocol Implementations via Conformance Testing Standardization". *IEEE International Conference on Communications ICC '90.* Atlanta, GA, 16-19 April 1990. New York; IEEE, 1990. 1374-1381.

Rose, Marshall T. *The Open Book: A practical perspective on OSI.* Englewood Cliffs; Prentice Hall, 1990.

Vermur, G.S. and Blik, H. "Interoperability Testing: Basis for the acceptance of communication systems (theory and practice)". (Sixth International Workshop on Practical Test Systems. IFIP TC6/WG6.1 Pau, France, 28-30 Sept. 1993) *IFIP Transactions C (Communication Systems)*, 1993. C-19:315-30.

Requirements for Accommodating Information Systems Information and Records Management Needs within the Proposals for a Government Information Locator Service (GILS) and its Z39.50 Application

David Bearman
Consultant
Archives & Museum Informatics
March 7, 1994

Requirements for Accommodating
Information Systems Information
and Records Management Needs
within the proposals for a
Government Information Locator Service (GILS)
and its Z39.50 Application

INTRODUCTION

As part of a project to study using Z39.50 in an application for the Government Information Locator Service (GILS), this contractor was hired to write a background paper outlining the needs of the archival and records management user communities.

The goal of this report is to "provide the research project with an increased understanding of the needs of the archival and records management communities regarding description of and access to Federal information resources and suggest enhancements or changes to the developing GILS Profile to support the requirements based on these needs."

Objectives include:

1) Outline the scope of the problem for identifying, describing, and accessing information about Federal information resources that are of interest to the archival and records management communities.

2) Describe the unique information needs of these communities.

3) Identify the functional requirements that can be derived from these information needs.

4) Suggest enhancements or changes to the emerging GILS Profile as documented in "Using Z39.50 in an Application for the Government Information Locator Service (GILS)" (current draft dated January 24, 1994)

I. ISSUES OF CONCERN TO THE ARCHIVAL COMMUNITY AND CITIZENS

Making and keeping records are essential to a democracy. They document how the government acted and why. By law, in the United States, these records may be obtained by citizens under the Freedom of Information Act (FOIA). Records pertaining to individuals will not be released in full under the Freedom of Information Act so as to protect personal privacy, but any citizen may obtain records pertaining to themselves under the terms of the Privacy Act which additionally gives them the right to correct any mis-information such records may contain.

The Federal Records Act requires agencies to make and keep records of the activity of the U.S. Federal Government for the period of their continuing value. The General Services Administration and the National Archives and Records Service share with the creating agency the responsibility for maintaining these records. Agencies are required to submit all records created in the course of business for scheduling by the National Archives; this involves identifying these records and determining the period of time for which they should be kept. Agencies may not destroy any records not authorized for destruction by an approved records schedule. After the period of active use by agencies has expired, records still being kept under the terms of an active schedule may be appraised and accessioned by the National Archives.

The system as described has a number of problems which are relevant to the GILS proposal:

A. Agencies have no simple way to report to the GSA and NARA when new recordkeeping systems are created in order to get a schedule and disposition authority.

B. Citizens have no comprehensive lis of records created by the Federal Government in the conduct of its business, despite the fact that all such records are required to be publicly accessible unless specifically exempted under limited clauses of the Freedom of Information and Privacy Acts. The right of citizens to know what records are being produced by their government, and, unless specific exemptions apply, to see those records on request is severely limited when they are unable to exercise their right because they do not know what government records exist. Citizens are not interested in where the records are currently held but in getting access to them, or to the information they contain.

C. The GSA and NARA have no way of using agency directories to records in order to fulfill their statutory responsibilities.

The General Services Administration has a records management program to assure that considered action is taken to retain or dispose of records created by the government in the course of its business. The National Archives and Records Administration must guarantee that records which are considered to have continuing value will be preserved and made accessible over time. Both agencies must necessarily know about all records created in the course of business and "schedule" or "appraise" such records. However neither agency has a complete database of records being created by the Federal Government.

The Government Information Locator Service is a possible solution:

The GILS proposal combines in a single service the ability

to meet both the operational requirements of records and archives management and the informational requirements of citizens. As such it is an important addition to the tools available to citizens of the United States and to the GSA and NARA, however it can only fulfill its intended function if it is implemented with some modifications to the proposed scope, data content and service definitions.

II. Functional Requirements

The problems identified in section I above lead to the identification of functional requirements for a Government Information Locator Service or similar directory to Federal information holdings.

1. Nothing short of a comprehensive information system on government information systems can satisfy the functional requirements of archives and records management or of citizens access to information. A comprehensive system would have to contain a record describing every series of Federal records as defined by current GSA and NARA regulations.

2. When recordkeeping system metadata is accessed, it must include the terms of the records schedule governing the records so that the continued existence of the records can be determined. General records schedules or approved special records schedules may be cited. The specific time period for which records must be retained under the terms of the schedule should be indicated. When no approved schedule yet exists, the system should permit the agency to request such a schedule.

3. Records created in the conduct of business are not typically assigned meaningful titles and do not have "subjects" in the sense that consciously authored bibliographic products do. Allowance should be made for archival practices which include assignment of constructed titles based on agency and program name, form of material, and span dates. Allowance also needs to be made for the use of function terms in place of subject terms in archival description.

4. There should not be redundant systems within the Federal government to which agencies must report records. The current requirement for agencies to fill out and submit NARA Form SF115 to request records scheduling could be dispensed with if GILS was accepted as a reporting mechanism.

5. End user access approaches presumed by the choice of Z39.50 services being supported are not explicitly identified. Since we have some knowledge of archival search strategies, a model could be developed based on prior research. Alternatively, the specification could incorporate recognition of the need to study use and modify choices. Such an approach should probably be incorporated into plans for the GILS anyway.

III. The Scope of GILS

In the January 22, 1994 draft document describing the Government Information Locator Service, the description of the scope of information systems to be reported to GILS is ambiguous. In order to have the GILS serve as a vehicle for informing citizens about government information available to them and as a tool for archives and records management, it is essential that the scope of the information systems to be included must be unambiguously stated as comprehensive.

It is recognized, of course, that the GILS Core will reference GILS agency components which may contain details of individual recordkeeping systems, but unless the service effectively acts as an all inclusive listing of government information, ancillary and alternative approaches will always be required at the cost of excessive redundant effort. For example, the National Archives will have to maintain its existing records scheduling system and records managers in agencies will need to create Agency records schedules listing essentially the same data but in a separate instrument. Citizens will still need to resort to the Freedom of Information Act to find out what information systems exist, and then to obtain information or records from them. Both the statement of GILS from the User Perspective and GILS from the Provider Perspective need to explicitly articulate the responsibility of Federal Agencies to assure that the GILS Core data links to agency-based locators and information services that in the aggregate identify all the recordkeeping systems of the agency. Therefore the proper specification of the Service should seek to realize the cost avoidance that would be possible if the system was comprehensive and accepted as the definition of what information was held by the Government.

In descriptions of the rationale for GILS, such as the Context section of the January 22 document, it should be clear that the intention of GILS is to not just to "help the public locate government information maintained by or for the agency" and "provide information describing how the public may gain access to agency information resources", but to assure that the public is informed of all information and records created by the government and explicitly assert that GILS is designed to satisfy the "fundamental requirement that Federal agencies maintain readily accessible inventories of their records and other information holdings." It is critical that agencies and citizens understand that the statement in 2.1 that "the public should be able to discover sources of publicly accessible information maintained throughout the U.S.Federal Government" refers to all information and records made or received in the course of business which are not explicitly exempted under FOIA or the Privacy Act.

IV. Data Content of GILS

The January 22, 1994 GILS draft, is much improved over prior drafts in meeting archives and records management requirements.

Under Mandatory Elements (4.3.1), the draft makes an assumption about the descriptive content of "Title" which seems unlikely to be satisfied by the titles given to most agency recordkeeping and information systems. I recommend renaming "Title" as "Descriptive Title/Label" and rewording the definition to read: "Since the descriptive title/label is intended for initial presentation to users independently of other elements, it must convey the most significant aspects of the referenced resource and provide sufficient information to allow users to make a decision on likely relevance. It should convey the most significant information available: in the case of information systems this should include the general topic are as well as reference to the specific subject; for recordkeeping systems it should reference the business function supported by the system, such as licensing, inspection, grants administration etc."

Under "Elements Mandatory for Information Systems" (4.3.2) the prose should be slightly revised and a data element should be added to document the records schedule.

1) Rephrase the introductory paragraph to read:

"The GILS Core includes a locator record for each Federal information system holding publicly accessible records or information."

or simply

"The GILS Core includes a locator record for each Federal information system."

The following additional data element is required, at a minimum, for every recordkeeping system: Scheduled Disposition. The values would be the period of retention expressed as months or years after creation of the record, or "not scheduled", "scheduled archival" or a prose description of the period of retention when not dated from creation of the record.

Other issues raised by the mandatory elements include:

Control Identifier: This should be consistent with the identifier used for other government reporting applications, such as reporting to NARA or GSA. If such identifiers have not been established already, perhaps NASA and GSA should commit themselves to using this identifier or providing a value that can be uniformly used.

Purpose: It would be better to state that the element "described why the information resource was created" rather than why it "is offered" since the reason for its being "offered" is that all government created or maintained information is publicly available.

V. Service Definitions for the GILS application:

The January 24 1994 draft "Using Z39.50 in an Application for the Government Information Locator Service (GILS)" is a excellent proposal for implementing GILS. The approach taken by the R&D project towards developing a consensus of stakeholders and specifying a standards-based implementation approach is exemplary.

1. Providing access to GILS via Z39.50 protocols is a good choice.

2. Without a more explicit statement by the developers of the assumptions they are making about the anticipated user population and its needs for the GILS, the configuration of tools proposed by the January 24 draft application cannot be validated. Specifically, it is difficult to determine if the decision not to support the Z39.50 Result-Set-Delete facility, Browse facility, Sort facility, access control facility, accounting/resource control facility, explain facility and extended services facility is appropriate.

Within the facilities that are proposed, several questions are still to be answered (or if answered, need to be made more explicit):

a) Why are the agencies with responsibility for implementing 44 U.S.C. not using the GILS Service as a means for satisfying agency obligations under this chapter?

> Under Core Requirements, Functions (4.1) in the GILS Document of January 22, 1994, it is stated that:
>
> "The GILS Core is designed to satisfy Federal agency responsibilities to maintain an inventory of their electronic information dissemination products as described in OMB Circular A-130. It should also be useful to agencies in improving agency responsiveness to FOIA requests. By including a record for each Federal information system holding publicly accessible data or information, the GILS Core thereby supports records management responsibilities of Federal agencies in reporting on agency information systems, codified in 44 U.S.C. Chapters 31 and 33. However maintaining in GILS a reference to the availability of an information product does not in itself satisfy all agency obligations under 44 U.S.C."

Ideally, for both public users and the GSA/NARA users, the agency GILS Core record reporting the inventory of agency information systems would have one consistent format and the locator records to which it pointed in the agency would be maintained and current. In this way, the GILS database would be an up-to-date view of agency record systems and their management.

b) Since agency locator records systems will themselves be represented in locators, and other agencies may construct locators that are "directories of directories", could a defined naming convention for such directories be suggested in the design or at least an assigned indexing term so that these could be uniformly identified? Such a convention should distinguish between "GILS Directory" records and records which have the title beginning with the term "Directory".

c) How will individuals navigate down to further levels of such locators? Using the available linkage in machine-readable form to connect to the information resource? There needs to be an executable, explicit link made in the local systems environment and this will have to be specified by GILS.

d) Why is there no specification for the attributes of Position, Truncation or Completeness?

e) Why is there no reference to Title or Abstract in Use Attributes?

f) Why is there no reference to equal in relation? Why not <?

Finally, as a comment on process, I would like to congratulate the developers for seeking opinion from a broad array of sources and note that the idea of a mechanism for testing interoperability is a good one. Opportunity for additional feedback would be valuable. I suggest using some statewide Information Locator Systems with Archives and Records Management input in the experiment. The States of Kentucky and New York have both established such locator based approaches to electronic records management.

Working Implementation Agreements for Open Systems Environment: Part 31 - Application Profile for the Government Information Locator Service (GILS) - Library Applications Special Interest Group

Output from the June 1994 Open Systems Environment Implementors' Workshop (OIW)

SIG Chair: **Ralph LeVan, Online Computer Library Center, Inc.**
SIG Editor: **Ralph LeVan, OCLC** Workshop Editor: **Brenda Gray, NIST**

Foreword

This part of the Working Implementation Agreements was prepared by the Library Applications Special Interest Group (LASIG) of the Open Systems Environment Implementors' Workshop (OIW). See Part 1 - Workshop Policies and Procedures in the "Draft Working Implementation Agreements" for the workshop charter.

Text in this part has been approved by the Plenary of the Workshop.

Future changes and additions to this version of these Implementor Agreements will be published as a new part. Deleted and replaced text will be shown as struck. New and replacement text will be shown as shaded.

Annex E (informative)

GILS Core Elements 33

Part 31: Application Profile for the Government Information Locator Service (GILS) - Library Applications SIG 1

0 Introduction 1

1 Scope and field of applications 2

2 Normative References 3

3 Definitions and terminology 4

4 Z39.50 Specifications for GILS 5
4.1 Version 5
4.2 GILS Objects 5
4.3 Communication Services 6
4.4 Z39.50 Services 6
4.4.1 Search 6
4.4.1.1 Attribute Set 6
4.4.1.2 Well-known Search 7
4.4.2 Retrieval 7
4.4.2.1 Schema 7
4.4.2.2 Element Sets Names 8
4.4.2.3 Record Syntaxes 8
4.5 Preferred Display Format for Use with SUTRS 9
4.6 Diagnostic Messages 9

5 Data Elements in the Locator Records 10

Annex A (informative)

GILS Attribute Set 11

Annex B (informative)

GILS Core Element to USMARC Mapping 15

Annex C (informative)

Preferred Display Format for GILS Records 22

Annex D (informative)

GILS Schema 24

Annex E (informative)

GILS Core Elements 40

List of Tables

Table 1 - Recognized and Supporte Combinations of GILS Attributes 11

Part 31: Application Profile for the Government Information Locator Service (GILS) - Library Applications SIG

0 Introduction

This document describes an application profile for the Government Information Locator Service (GILS). The GILS Profile includes not only the specifications for ANSI/NISO Z39.50, the American National Standard for Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (National Information Standards Organization, 1992) in the application but also other aspects of a GILS conformant server that are outside the scope of Z39.50. The GILS Profile provides the specifications for the overall GILS application relating to the GILS Core, which is a subset of all GILS Locator Records, and completely specifies the use of Z39.50 in this application.

Background

The GILS is a response to the need for users to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources. Christian (1994) is the authoritative document providing an overview of GILS, its objectives, service requirements, and core requirements. According to Christian (1994), the GILS is an overall service and includes information and technology components as well as policy, regulation, people, etc. The GILS is intended to help the public locate and access public information throughout the U.S. government.

The current GILS initiative builds upon a previous study, Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators (McClure, Ryan & Moen, 1992). That study, which was conducted for the Office of Management and Budget, the National Archives and Records Administration, and the General Services Administration, recommended that each agency establish a network-accessible locator that describes its information resources. The study also recommended that agencies use Z39.50 as the appropriate information retrieval protocol to achieve a distributed, standards-based Government Information Locator Service.

The development work of the GILS Profile is documented in Using Z39.50 in an Application for the Government Information Locator Service (GILS) (McClure & Moen, 1994). The GILS Profil resulted from the work of a group comprising experts in Z39.50 implementations, system implementations, and information organization, and representatives of Federal agencies. The specifications included in the GILS Profile reflect the consensus of this group and input from a range of stakeholders.

1 Scope and field of applications

The GILS Profile fully specifies the use of ANSI/NISO Z39.50 by the GILS. In addition, the GILS Profile provides the specifications for the overall GILS application relating to the GILS Core including other aspects of GILS conformant servers that are outside the scope of Z39.50.

This version of the GILS Profile focuses on requirements for a GILS server operating in the Internet environment. GILS clients will be able to interconnect with any GILS server, and these clients will behave in a manner which allows interoperability with the GILS server. Clients that support Z39.50 but do not implement the GILS Profile will be able to access GILS records with less than full GILS functionality.

The GILS Profile addresses many aspects of the GILS (e.g., intersystem interactions and information interchange) but does not specify user interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality.

Field of Application

The GILS Profile supports search and retrieval of GILS Locator Records contained in GILS servers by users in the Internet environment.

The GILS Profile will be used by developers of GILS servers. It will also be used by client developers to understand expected behaviors of GILS servers. A GILS server accessed using Z39.50 in the Internet environment acts primarily as a pointer to information resources. Some of these information resources pointed to by GILS Locator Records, as well as the GILS server itself, may be available electronically through other communications protocols including the common Internet protocols that facilitate electronic information transfer such as remote login (Telnet), File Transfer Protocol (FTP), and electronic mail (SMTP/MIME). The use of these protocols or other communications paths is outside the scope of the GILS Profile.

Once connected to a GILS server, users supported by appropriate clients that understand the GILS Profile may navigate through single or multiple servers. GILS servers will support searching (i.e., accept a search query and return a result set or diagnostic messages) and may support browsing (i.e., accept a well-known search query and return a list of Locator Records in brief display format). Although the GILS Profile addresses GILS servers only, it is understood that clients have roles in the execution of these activities (e.g., browsing is also a client function in the sense of how it interprets and presents GILS data).

2 Normative References

The following list contains documents that contain provisions which, through reference in this text, constitute provisions of the GILS Profile. At the time of this publication, the editions indicated were valid. All documents are subject to revision, and parties to agreements based on this Profile are warned against automatically applying any more recent editions of the documents listed below, since the nature of references made by the Profile to such documents, is that they may be specific to a particular edition. In addition, this list contains other documents that can be consulted for further information, background, etc.

[1] American National Standards Institute. (1985). American National Standard 39.2-1985 Bibliographic Information Interchange. New York: American National Standards Institute.

[2] Christian, Eliot. (1994, May 2). Government Information Locator Service (GILS): Report Information Infrastructure Task Force. Available on the Fedworld electronic bulletin board (703-321-8020) or by anonymous FTP (File Transfer Protocol) via the Internet at 130.11.48.107 as /pub/gils.doc (Microsoft Word for Windows format) or /pub/gils.txt (ASCII text format).

[3] Lynch, Clifford A. (1994, April 30). "Using the Z39.50 Information Retrieval Protocol in the Internet Environment" [Draft RFC for Z39.50 over TCP/IP].

[4] McClure, Charles R. & Moen, William E. (1994, may 7). Using Z39.50 in an Application for the Government Information Locator Service (GILS). Available via anonymous FTP at <ericir.syr.edu> as /USGS/profile_background.doc.ps (Postscript format) and as /USGS/profile_background.doc.txt (ASCII text format).

[5] McClure, Charles R., Ryan, Joe & Moen, William E. Moen, (1992). Identifying and Describing Federal Information Inventory/Locator Systems: Design for Networked-Based Locators. 2 Vols. Bethesda, MD: National Audio Visual Center [Available from ERIC, document no. ED349031].

[6] National Information standards Organization, (1992). ANSI/NISO Z39.50-1992, Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. Gaithersburg, MD: NISO Press.

[7] National Institute of Standards and Technology. (1992). FIPS No. 173, Spatial Data Transfer Standard (August 28, 1992). Gaithersburg, MD: National Institute of Standards and Technology.

[8] Office of Management and Budget. (1993). Circular No. A-130, "Management of Federal Information Resources" (58 F.R. 36068, July 2, 1993).

[9] Open Systems Environment Implementors Workshop/Special Interest Group on Library Applications (OIW/SIGLA), (1993). OIW/SIGLA Document #1: Using Z39.50-1992 Directly over TCP.

[10] RFC 1521, MIME (Multipurpose Internet Mail Extensions) Part One: Mechanisms for Specifying and Describing the Format of Internet Message Bodies.

[11] RFC 1522, MIME (Multipurpose Internet Mail Extensions) Part Two: Message Header Extensions for Non-ASCII Text.

[12] Uniform Resource Locators (URL): A Unifying Syntax for the Expression of Names and Addresses of Objects on the Network, (1993, October). [Internet Draft]. The latest URL draft is: <url:ftp://info.cern.ch/pub/www/doc/url7a.txt>.

[13] Uniform Resource Names, (1993 October). [Internet Draft]. The latest URN draft is: <url:ftp://ds.internic.net/internet-drafts/draft-ietf-uri-resource-names-01.txt>.

[14] USMARC Format for Bibliographic Data. Washington, DC: Library of Congress, Cataloging Distribution Service.

3 Definitions and terminology

For purposes of this Profile, the following definitions apply.

Client: An initiating application. This application includes the Z39.50 origin.

Electronic Information Resource: Information resources that are maintained in electronic, digital format and may be accessed, searched, or retrieved via electronic networks or other electronic data processing technologies (e.g., CD-ROM).

GILS Core: A subset of all GILS Locator Records which describe information resources maintained by the U.S. Federal government and comply with the defined GILS Core Elements and are mutually accessible through interconnected electronic network facilities without charge to the direct user.

Government Information: Information created, collected, processed, disseminated, or disposed of by or for the Federal government.

Government Information Locator Service (GILS): A decentralized collection of locators and associated information services used by the public either directly or through intermediaries to find public information throughout the U.S. Federal government.

Information Resource: Includes both government information and information technology.

Interoperability: A condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.

Locator Record: A collection of related data elements describing an information resource, the information available in the resource, and how to obtain the information.

Mandatory: An element in a GILS Core Locator Record that must have a value provided by the record source. The GILS Profile does not specify which elements must be present from the perspective of GILS servers.

Origin: The part of a client application that initiates a Z39.50 association and is the source of requests during the association.

Profile: The statement of a function(s) and the environment within which it is used, in terms of a set of one or more standards, and where applicable, identification of chosen classes, subsets, options, and parameters

of those standards. A set of implementor agreements providing guidance in applying a standard interoperably in a specific limited context.

Registered Object: An object that is identified by a name-to-thing relationship in which the name is recorded by a registration authority to ensure that the names can be used unambiguously.

Server: An application that responds to an initiating application (i.e., a client). The application that includes the Z39.50 target.

Target: The part of an server application that accepts a Z39.50 association.

Uniform Resource Identifier (URI): A set of related standards for encoding resource location and identification information for electronic and other objects. Examples include Uniform Resource Locators (URLs) and Uniform Resource Names (URNs).

USMARC: An implementation of ANSI/NISO Z39.2, the American National Standard for Bibliographic Information Interchange. The USMARC format documents contain the definitions and content designators for the fields that are to be carried in records structured according to Z39.2. GILS records in USMARC format contain fields defined in USMARC Format for Bibliographic Data. This documentation is published by the Library of Congress.

4 Z39.50 Specifications for GILS

This section details the required services available from Z39.50, describes an Attribute Set for searching and four Element Sets Names by which the server presents some or all the elements, defined in the Schema, of the Locator Records, and prescribes the Record Syntaxes to be supported by GILS servers for the transfer of Locator Records.

4.1 Version

GILS clients and servers support Z39.50 Version 2 as specified in Z39.50-1994. GILS requires support of various objects, some of which are not defined in Z39.50-1992. These are listed in 4.2.

4.2 GILS Objects

The following object identifier (OID) is assigned to the Z39.50 standard:

{iso (1) member-body (2) US (840) ANSI-standard Z39.50 (10003)}

This OID is abbreviated as: ANSI-standard-Z39.50.

Several object classes are assigned at the level immediately subordinate to ANSI-standard-Z39.50, including:

- 3 = attribute set definitions;

- 4 = diagnostic definitions;
- 5 = record syntax definitions;
- 13 = database schema definitions;
- 14 = tagSet definitions.

GILS requires support of the following objects:

- GILS attribute set: {ANSI-standard-Z39.50 3 3};
- bib1 diagnostic set: {ANSI-standard-Z39.50 4 1};
- USMARC record syntax: {ANSI-standard-Z39.50 5 10};
- SUTRS record syntax: {ANSI-standard-Z39.50 5 101};
- GRS-1 record syntax: {ANSI-standard-Z39.50 5 105};
- GILS schema: {ANSI-standard-Z39.50 13 2};
- tagSet-M: {ANSI-standard-Z39.50 14 1};
- tagSet-G: {ANSI-standard-Z39.50 14 2}.

4.3 Communication Services

When Transmission Control Protocol (TCP) is used as the transport service, the specification for use of TCP is found in OIW/SIGLA Document #1, "Using Z39.50-1992 Directly over TCP." The use of other communication services is not yet defined.

4.4 Z39.50 Services

There are three Z39.50 (Version 2) services that are required for conformance: Init, Search, and Present. No additional services are required for conformance to the GILS Profile. Other Z39.50 services, however, may be provided optionally by servers and used by clients.

Standard Z39.50 Init Service negotiation procedures control the use of all services.

4.4.1 Search

The GILS application will support Z39.50 Type 1 queries which are general purpose Boolean query structures.

4.4.1.1 Attribute Set

The GILS Attribute Set is a superset of the Bib-1 Attribute set and consists of all Bib-1 Attributes and additional Use Attributes that are defined for GILS elements (see Annex A for the GILS Use Attributes). These newly defined GILS Use Attributes are well-known and correspond in name and semantics to the elements in the GILS Schema. The GILS Attribute Set is a registered object.

GILS servers must support a limited number of GILS Attributes. The required GILS Attributes are: (Note: GILS Use Attribute Name is listed followed by the GILS Use Attribute Number and the corresponding GILS Core Element Name):

- Use Attributes: Local Number (12; Local Control Number); Author-name corporate (1005; Originator); Date/Time Last Modified (1012; Date of Last Modification); Record Source (1019; Record Source); Distributor Name (2001; Distributor Name); Index Terms – Controlled (2002; Index Terms – Controlled); Local Subject Index (29; Local Subject Term); Any (1016)

- Structure: Word (2), URx (104), Date (5), Word List (6)

- Relation: Greater than (5), Equal (3).

GILS servers should never return any of these four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes the combinations of required GILS Attributes listed in Table 1 in Annex A.

4.4.1.2 Well-known Search

To provide support for browsing GILS Locator Records, there is a well-known search consisting of the GILS Attribute Set Use Attribute: Local Number; Structure Attribute: URL; and a term of zero length. GILS servers that support browsing of records will create a result set of one or more GILS Locator Records that provide the necessary information to allow clients to offer menu-like displays of GILS Locator Records or other information and information resources.

The "Browse" in the GILS context involves only the Search and Present Services of Z39.50. "Browse" is used informally in the GILS Profile, and it is not related nor should it be confused with the Browse Facility or Scan Service of Z39.50.

BEST COPY AVAILABLE

4.4.2 Retrieval

This section describes the components and procedures used by Z39.50 to return records in response to a query.

4.4.2.1 Schema

The GILS Profile specifies a GILS Schema (see Annex D for the Schema). The GILS Schema is a registered object. A schema in Z39.50 can be modified and may evolve over time, and it is reasonable to expect the GILS Schema will evolve.

The GILS Schema uses elements from tagSet-M and tagSet-G and defines in the GILS tagSet additional elements as necessary. The GILS Profile specifies tagTypes to identify tagSet-M elements (tagType = 1), tagSet-G elements (tagType = 2), and the elements defined by the GILS tagSet (tagType = 4). Another tagType (tagType = 3) is used to identify arbitary string tags for locally defined elements.

The GILS tagSet element numbering begins with number 1. Elements can be nested and the tagging notation (i.e., the tag path) will reflect the nesting.

All well-known GILS Schema elements have assigned numeric tags. String-tags (i.e., text) may be used in the GILS Schema to label those elements that are not well-known (i.e., locally defined).

4.4.2.2 Element Sets Names

GILS servers will support four Element Sets Names. GILS servers will interpret the use of the Element Set Names required by the GILS Profile to identify the following elements from the GILS Schema:

- The primitive element set name "B" contains at least: title, control identifier, originator, and local control number;
- The primitive element set name "G" contains: all B Element Set elements and Cross Reference;
- The primitive element set name "W" contains: all B Element Set elements and bodyOfDisplay;
- The primitive element set name "F" contains all elements available in the record.

The server should include in a retrieved record all of the elements specified by the element set name for which there is data available in the database record and which can be encoded in the requested record syntax (e.g., some types of locally defined binary data may not be encodable in a USMARC or SUTRS record).

4.4.2.3 Record Syntaxes

GILS servers are required to support the following three record syntaxes:

- USMARC – an implementation of ANSI/NISO Z39.2 and maintained by the Library of Congress;
- Generic Record Syntax (GRS-1) – defined in Z39.50;
- Simple Unstructured Text Record Syntax (SUTRS) – defined in Z39.50.

Annex B contains a mapping of Core Elements to USMARC for use in the USMARC record syntax. However, since the data transformation is not fully reversible and requires interpretation, the record source is responsible for encoding the USMARC record(s).

The data in GILS Locator Records do not always map clearly into USMARC records, particularly when agencies add their own locally defined fields to the GILS Locator Record. This means that construction of USMARC records is subject to local interpretation. Therefore, GILS Locator Records in USMARC format obtained from other than the original record source should be considered non-definitive. The original source of the GILS Locator Record can be identified by examining the Original Control Identifier field of the record.

For interchange, GRS-1 records are to be treated as the complete and canonical representation; SUTRS and USMARC should be viewed as derivative records from these canonical representations and as such are not as complete or precise.

4.5 Preferred Display Format for Use with SUTRS

The GILS Profile recommends a preferred display format for SUTRS records (see Annex C for the recommended display format). For the SUTRS records, formatting instructions for a preferred display format is a concern of the server.

When the target transfers a GILS record using the SUTRS record syntax, it will encode the GILS record formatted according to the preferred display format, so that the client may present the record directly, without processing. For SUTRS, however, the client should not expect to be able to parse the record to obtain any individual GILS elements.

When the client presents a GILS record formatted by the server using the USMARC or GRS record syntax, it is recommended that the client consider the SUTRS suggested display layout in formatting the received record for presentation to the human end user.

4.6 Diagnostic Messages

The GILS application will use Diagnostic Set Bib-1.

5 Data Elements in the Locator Records

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The GILS Core Elements are defined in Annex E.

Annex A (informative)

GILS Attribute Set

The GILS Attribute Set is a superset of the Bib-1 Attribute Set and consists of all Bib-1 Attributes and the additional Use Attributes listed below. Additional Use Attributes that cannot be mapped to Bib-1 Use Attributes are numbered from 2000 through 2999. These are well-known Use Attributes.

GILS servers should never return any of these four diagnostic messages: "Unsupported Use Attribute," "Unsupported Structure Attribute," "Unsupported Position Attribute," or "Unsupported Attribute Type" when a query includes the combinations of GILS Attributes listed in table 1. An "X" in the table means that GILS servers will recognize and support this combination of Attributes.

Table 1 - Recognized and Supported Combinations of GILS Attributes

USE	WORD	URx	DATE	WORD LIST	GREATER THAN	EQUAL
Local Number	X	X		X		X
Author-name corporate	X			X		X
Date/Time Last Modified			X		X	X
Record Source	X			X		X
Distributor Name	X			X		X
Index Term - Controlled	X			X		X
Local Subject Index	X			X		X
Any	X			X		X

As stated in 4.3.1.1, GILS servers are required to support a minimal set of Use Attributes. These are listed first. In the cases where a Bib-1 Use Attribute's Name is used, the corresponding GILS Core Element name appears in parentheses.

Required GILS Use Attributes	
Use #	GILS Attribute Name
12	Local Number (Local Control Number)
29	Local Subject Index (Local Subject Term)
1005	Author-name corporate (Originator)
1012	Date/Time Last Modified (Date of Last Modification)
1016	Any
1019	Record Source
2001	Distributor Name
2002	Index Terms – Controlled

Available GILS Use Attributes	
USE #	GILS Attribute Name
4	Title
1007	Identifier - Standard (Control Identifier)
62	Abstract
2003	Purpose
2004	Access Constraints
2005	Use Constraints
2006	Distributor Organization
2007	Distributor Street Address
2008	Distributor City
2008	Distributor State
2010	Distributor Zip Code
2011	Distributor Country
2012	Distributor Network Address
2013	Distributor Hours of Service
2014	Distributor Telephone
2015	Distributor Fax
2016	Available Resource Description
2017	Available Order Process
2018	Available Technical Prerequisites
2019	Available Time Period – Structured
2020	Available Time Period – Textual
2021	Available Linkage
2022	Available Linkage Type
2023	Contact Name
2024	Contact Organization
2025	Contact Street Address
2026	Contact City
2027	Contact State

2028	Contact Zip Code
2029	Contact Country
2030	Contact Network Address
2031	Contact Hours of Service
2032	Contact Telephone
2033	Contact Fax
2034	Agency Program
2035	Sources of Data
2036	Thesaurus
2037	Methodology
2038	Bounding Rectangle – Western-most
2039	Bounding Rectangle – Eastern-most
2040	Bounding Rectangle – Northern-most
2041	Bounding Rectangle – Southern-most
2042	Geographic Keyword Name
2043	Geographic Keyword Type
2044	Time Period - Structured
2045	Time Period - Textual
2046	Cross Reference Title
2047	Cross Reference Linkage
2048	Cross Reference Type
2049	Original Control Identifier
2050	Supplemental Information

Annex B (informative)

GILS Core Element to USMARC Mapping

This annex provides a mapping from GILS Core Elements to USMARC for use by the record source and GILS servers. Some of these data elements consist of two or more subelements, and this relationship is noted by the indentation.

Implementors should consult the authoritative documentation on USMARC found in USMARC Format for Bibliographic Data. The document is available from the Cataloging Distribution Service at the Library of Congress. A full description of the USMARC fields and available subfields within each field is in that document.

For some elements new USMARC fields and/or subfields may be incorporated into the USMARC format. New fields and/or subfields in the process of being considered for inclusion in USMARC are noted.

In cases where the 500 Note field is repeated to carry separate GILS Core Elements, the name of the GILS Core Element will be included and precede the data content for that field. A colon will separate the GILS Data Element name from the rest of the content in the field. For example, 500 Purpose: [data for this field]; 500 Agency Program: [data for this field]. Each such GILS Core Element should be carried in separate, repeating 500 fields.

In addition to the variable length fields listed in the mapping, a USMARC record will also include a Leader and field 008: Fixed-Length Data Elements. Certain character positions in each of these fixed length fields of a USMARC record will need to be coded specifically for GILS. In addtion, USMARC records for GILS will include a code in the 042: Authentication Code to identify these USMARC records psecifically as GILS Locator Reocrds. The following sugest values for these fields (or paths of these fields):

Leader: A fixed field comprising the first 24 character positions (00-23) of each record that provides information for the processing of the record. For GILS records, the following character position is specifically relevant:

Character Position: 18 -- Descriptive cataloging form: Value: # [i.e., blank] (Non-ISBD) to indicate when International Standard Bibliographic Description is not followed.

008 Fixed Length Data Elements: Forty character positions (00-39) containing positionally-defined data elements that provide coded information about the record as a whole or about special bibliographic aspects of the item being cataloged. For GILS records that describe electronic information resources, the following character position is specifically relevant:

Character Postion: 26 – Type of computer file
Values: a (Numeric data)
b (Computer program)
c (Representational)
d (Document)
e (Bibliographic data)
f (Font)
g (Game)
h (Sound)
i (Online system or service) [new code proposed]
m (Combination)
u (Unknown)
z (Other)

042 Authentication Code:
Value: gils [new code proposed]

GILS Data Elements and Corresponding USMARC Tags	
GILS Data Element	USMARC Tag
Title	245$a
Control Identifier	001
Abstract	520
Purpose	500
Originator	710$a
Access Constraints	506
Use Constraints	540
Distributor	
Distributor Name	270$p [proposed field]
Distributor Organization	270$p [proposed field]
Distributor Street Address	270$a [proposed field]
Distributor City	270$b [proposed field]
Distributor State	270$c [proposed field]
Distributor Zip Code	270$e [proposed field]
Distributor Country	270$d [proposed field]
Distributor Network Address	270$m [proposed field]
Distributor Hours of Service	270$a [proposed field]
Distributor Telephone	270$k [proposed field]
Distributor Fax	270$1 [proposed field]
Available Resource Description	037$f
Available Order Process	037$c
Available Technical Prerequisites	538
Available Time Period – Structured	045$c

BEST COPY AVAILABLE

Available Time Period – Textual	037$n [proposed field]) (for non-electronic resource 856$z (for electronic resource)
Available Linkage	856$u
Available Linkage Type	856 1st indicator/856$2
Point of Contact	856$m (for electronic resources)
Contact Name	270$p [proposed field]
Contact Organization	270$p [proposed field] 535
Contact Street Address	270$a [proposed field] (for non-electronic resources)
Contact City	270$b [proposed field]
Contact State	270$c [proposed field]
Contact Zip Code	270$e [proposed field]
Contact Country	270$d [proposed field]
Contact Network Address	270$m [proposed field]
Contact Hours of Service	301$a [proposed field]
Contact Telephone	270$k [proposed field]
Contact Fax	270$1 [proposed field]
Record Source	040
Date Last modified	005
Agency Program	500
Sources of Data	537 [proposed field]
Index Terms – Controlled	650
Thesaurus	650 1st indicator/ 650$2
Local Subject Term	653$a

Methodology	567
Spatial Reference	
Bounding Rectangle	255$c
Western-most	034$d
Eastern-most	034$e
Northern-most	034$f
Southern-most	034$g
Geographic Name	
Geographic Keyword Name	651
Geographic Keyword Type	655
Time Period – Structured	045$c
Time Period – Textual	513
Cross Reference Title	787$t
Cross Reference Linkage	787$w
Cross Reference Type	856 1st indicator/856$2
Original control identifier	035
Supplemental information	500

USMARC Tags and Field Names (from USMARC Format for Bibliographic Data)		
USMARC Tag	Subfield	Field Name
001		Control Number
005		Date and Time of Latest Transaction
034		Coded Cartographic Mathematical Data
	$d	Coordinates - westernmost longitude
	$e	Coordinates – easternmost longitude
	$f	Coordinates – northernmost latitude
	$g	Coordinates – southernmost latitude
035		System Control Number
037		Source of Acquisition
	$b	Source of stock number/acquisition
	$c	Terms of availability
	$f	Form of issue
	$n	Note (proposed)
040		Cataloging Source
042		Authentication Code
245		Title Statement
	$a	Title
255		Cartographic Mathematical Data
	$c	Statement of coordinates
270	$a	Address
270	$b	City
270	$c	State or province
270	$d	Country
270	$e	Postal code
270	$k	Telephone number
270	$1	Fax number
270	$m	Electronic mail address
270	$p	Contact person

301	$a	Hours
500		General Note
506		Restrictions on Access Note
513		Type of Report and Period Covered Note
520		Summary, Etc. Note
537		Source of Data Note [proposed]
538		System Details Note
540		Terms Governing Use and Reproduction Note
567		Methodology Note
650		Subject Added Entry – Topical Term
1st indicator		Level of subject
	$2	Source of heading or term
651		Subject Added Entry – Geographic Name
653		Index Term – Uncontrolled
	$a	Uncontrolled term
655		Index Term – Genre/Form
710		Added Entry – Corporate Name
	$a	Corporate name or jurisdiction name as entry element
787		Nonspecific Relationship Entry
	$t	Title
	$w	Record Control Number
856		Electronic Location and Access
1st indicator		Access method
	$m	Contact for access assistance
	$u	Uniform Resource Locator
	$z	Nonpublic note
	$2	Source of access

Annex C (informative)

Preferred Display Format for GILS Records

GILS servers will transfer records in three record syntaxes:

- USMARC
- Generic Record Syntax (GRS)
- Simple Unstructured Text Record Syntax (SUTRS).

In SUTRS, the formatting of the record contents is handled by the server, and the client receives a record devoid of structure. In USMARC and GRS, the record, whose structure is defined by the record syntax, is passed from the target to an orgin, and the client software has more flexibility in processing the record contents for display.

The recommended guidelines in this Annex describe how records should be displayed, whether formatted by the server or the client (but this does not preclude display formats in addition to the Preferred Display Format).

Record Organization:

The record should be organized so that the elements first viewed by the user provide adequate information to either choose or eliminate the record from further consideration. These elements are: Title, Originator, Controlled Vocabulary, Local Subject Index and Abstract.

Next in the order of presentation are elements that give detailed information about the information resource being described: Spatial Reference, Time Period, Availability, Sources of Data, Methodology, Access Constraints, Use Constraints, Point of Contact, and Supplemental Information.

The elements describing the reason for the existence of the data are next: Purpose and Agency Program.

Related information resources are listed next in the element: Cross Reference

The final elements provide bibliographic control information: Control Identifier, Record Source, and Date of Last Modification.

General Instructions for Formatting Full Element Set Name Records:

All displayable elements are to be labelled with the full title of the field followed by a colon. Label mnemonics should only be used in situations where the user can ask for an explanation of the mnemonic. Mnemonics should not be used in SUTRS records, since it should be assumed that the client knows nothing about the server and is incapable of interpreting the mnemonics.

The subelements of constructed elements (i.e., locally defined fields, Availability, Spatial Reference, etc.) should be indented to reflect their association and structure within a well-structured element. Labels on subelements can eliminate the redundant leading parts (e.g., the word Available on the Availability

subelements).

In the Controlled Vocabulary element, the Thesaurus subelement can be presented in parentheses, followed by the Index Terms. Multiple Index Terms should be separated by a semi-colon and a space (e.g., Controlled Vocabulary (MeSH): Kidney; Kidney Disease). Alternatively, the Thesaurus and Index Terms can be indented under the Controlled Vocabulary label, as is done with the other well-structured fields. Local Subject Terms should be separated by a semi-colon and a space.

Display Format for Brief Element Set Name Records:

Brief Records consist of the Title, Control Identifier, Originator, and Local Control Number fields. For display purposes, the Control Identifier and Local Control Number can be omitted. Brief Records may be formatted to fit on a single line. This may require that one or both of the displayed fields will be truncated. Truncation can be indicated with ellipsis (...).

Display Format for G Element Set Name Records:

G Records consist of Brief Record elements and additionally, the Cross Reference element. For display purposes, the guidelines for Full Records should be followed.

Annex D (informative)

GILS Schema

The GILS Schema describes and defines tagSets and an Abstract Record Structure used with the Generic Record Syntax (GRS). The GILS Schema defines a GILS tagSet that associates a numeric tag with one or more GILS Core Elements.

Some GILS Core elements correspond to tags already defined in tagSetM and tagSet-G, these tags are used to identify GILS Core elements in the Abstract Reocrd Structure. When the tagType is 1, the tag value is from tagSet-M. When the tagType is 2, the tag value is from tagSet-G. When the tagType is 3, the tag value is an arbitrary string tag. When the tagType is 4, the tag value is from the GILS tagSet.

There are two general classes of schema elements in the GILS Schema:

1) Primitive -- these elements cannot have locally defined subelements;

2) Constructed -- these elements have one or more subelements any of which may be well-defined or target-defined; in the latter case, these locally defined subelements are identified with string tags.

This Annex first presents first the GILS tagSet that identifies the element, its unique tag, and a recommended datatype . This is followed by the GILS Abstract Record Structure that shows the full tag path for each element.

GILS tagSet		
Tag	Element	Recommended Datatype
1	ControlIdentifier	InternationalString
2	streetAddress	InternationalString
3	City	InternationalString
4	state	InternationalString
5	zipcode	InternationalString
6	hoursOfService	InternationalString
7	resourceDescription	InternationalString
8	technicalPrerequisites	InternationalString
9	westernMost	intUnit
10	easternMost	intUnit
11	northernMost	intUnit
12	southernMost	intUnit
13	geographicKeywordName	InternationalString
14	geographicKeywordType	InternationalString
15	timePeriodStructured	GeneralizedTime
16	timePeriodTextual	InternationalString
17	linkage	InternationalString
18	linkageType	InternationalString
19	recordSource	InternationalString
20	controlledTerm	InternationalString
21	thesaurus	InternationalString
22	localSubjectTerm	InternationalString
23	originalControlIdentifier	InternationalString

NOTE - The element "wellKnown" from tagSet-M (1,19) and referred to below has the following definition:

When an element is defined to be "structured into locally defined elements," the target may use this tag (i.e., wellKnown) in lieu of, or along with, locally defined tags. For example, an element named "title" might be described to be "locally structured." The target might present the element structured into the following subelements: "wellKnown," "spineTitle,"

and "variantTitle," where the latter two tags are target defined. In this case, "wellKnown" is assumed to mean "title."

GILS tagSET		
Tag	Element	Recommended Datatype
50	title	Constructed as follows–
	This element may include the element wellKnown and may also include locally defined elements.	
51	purpose	Constructed as follows–
	This element may include the element wellKnown and may also include locally defined elements.	
52	originator	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
53	accessConstraints	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
54	useConstraints	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
55	orderProcess	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
56	agencyProgram	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
57	sourcesOfData	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
58	methodology	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
59	supplementalInformation	Constructed as follows –
	This element may include the element wellKnown and may also include localy defined elements.	
70	availability	Constructed as follows –
	This element may include any of the following as well as locally defined elements: distributor, resourceDescription, orderProcess, technicalPrerequisites, timePeriod, linkage, linkageType.	

BEST COPY AVAILABLE

71	spatialReference	Constructed as follows –
	This element may include any of the following as well as locally defined elements: boundingRectange, geographicName.	
90	distributor	Constructed as follows –
	This element may include any of the following as well as locally defined elements: name, organization, streetAddress, city, state, zipCode, country, networkAddress, hoursOfService, phoneNumber, faxNumber.	
91	boundingRectangle	Constructed as follows –
	This element may include any of the following as well as locally defined elements: westernMost, easternMost, northernMost, southernMost.	
92	geographicName	Constructed as follows –
	This element may include any of the following as well as locally defined elements: geographicKeywordName, georgraphicKeywordType.	
93	timePeriod	Constructed as follows –
	This element may include any of the following as well as locally defined elements: timePeriodStructured, timePeriodTextual.	
94	pointOfContact	Constructed as follows –
	This element may include any of the following as well as locally defined elements: name, organization, streetAddress, city, state, zipCode, country, networkAddress, hoursOfService, phoneNumber, faxNumber.	
95	controlledVocabulary	Constructed as follows –
	This element may include any of the following as well as locally defined elements: indexTermsControlled, thesaurus.	
96	indexTermsControlled	Constructed as follows –
	This element may include any of the following as well as locally defined elements: controlledTerm.	
97	localSubjectIndex	Constructed as follows –
	This element may include any of the following as well as locally defined elements: localSubjectTerm.	
98	crossReference	Constructed as follows–
	This element may include any of the following as well as locally defined elements: title, linkage, linkageType.	

GILS Abstract Record Structure

NOTE - The element "bodyOfDisplay" in tagSet-G(2,9) may be used by the target to combine into this single element (i.e., bodyOfDisplay) one or more of the elements from the following abstract record structure into a display format.

Tag path	Element	Mandatory	Repeatable?
(1,10)	rank	N	N
(1,12)	url	N	N
(1,14)	local control number	Y	N
(1,16)	dateOfLastModification	Y	N
(4,50)	title	Y	N
(4,1)	controlIdetifier	Y	N
(2,6)	abstract	Y	N
(4,51)	purpose	Y	N
(4,52)	originator	Y	N
(4,53)	accessConstraints	Y	N
(4,54)	useConstraints	Y	N
(4,70)	availability	Y	Y
(4,70)/(4,90)	distributor	Y	N
(4,70)/(4,90)/(2,7)	distributorName	Y	N
(4,70/(4,90)/(2,10)	distributorOrganization	Y	N
(4,70/(4,90)/(4,2)	distributorStreetAddress	Y	N
(4,70/(4,90)/(4,3)	distributorCity	Y	N
(4,70/(4,90)/(4,4)	distributorState	Y	N
(4,70/(4,90)/(4,5)	distributorZipCode	Y	N
(4,70/(4,90)/(2,16)	distributorCountry	Y	N
(4,70/(4,90)/(2,12)	distributorNetworkAddress	Y	Y
(4,70/(4,90)/(4.6)	distributorHoursofService	Y	Y
(4,70/(4,90)/(2,14)	distributorPhoneNumber	Y	Y
(4,70/(4,90)/(2,15)	distributorFaxNumber	Y	Y
(4,70/(4,7)	resourceDescription	N	N
(4,70/(4,55)	orderProcess	Y	N
(4,70/(4,8)	technicalPrerequisites	N	N
(4,70/(4,93)	timePeriod	N	Y
(4,70/(4,93)/(4,15)	timePeriodStructured	N	Y

(4,70/(4,93)/(4,16)	timePeriodTextual	N	Y
(4,70/(4,17)	linkage	N	N
(4,70/(4,18)	linkageType	N	N
(4,94)	pointOfContact	Y	N
(4,94)/(2,7)	contactName	Y	N
(4,94)/(2,10)	contactOrganization	Y	N
(4,94)/(4,2)	ContactStreetAddress	Y	N
(4,94)/(4,3)	ContactCity	Y	N
(4,94)/(4,4)	ContactState	Y	N
(4,94)/(4,5)	ContactZipCode	Y	N
(4,94)/(2,16)	ContactCountry	Y	N
(4,94)/(2,12)	ContactNetworkAddress	Y	Y
(4,94)/(4,6)	ContactHoursofService	Y	Y
(4,94)/(2,14)	ContactPhoneNumber	Y	Y
(4,94)/(2,15)	ContactFaxNumber	Y	Y
(4,19)	recordSource	Y	N
(4,56)	agencyProgram	N	N
(4,57)	sourcesOfData	N	N
(4,95)	controlledVocabulary	N	Y
(4,95)/(4,96)	indexTermsControlled	Y	N
(4,95)/(4,96)/(4,20)	controlledTerm	Y	Y
4,95)/(4,21)	thesaurus	Y	N
(4,97)	localSubjectIndex	N	N
(4,97)/(4,22)	localSubjectTerm	Y	Y
(4,58)	methodology	N	N
(4,71)	spatialReference	N	N
(4,71)/(4,91)	boundingRectangle	N	N
(4,71)/(4,91)/(4,9)	westernMost	N	N
(4,71)/(4,91)/(4,10)	easternMost	N	N
(4,71)/(4,91)/(4,11)	northernMost	N	N

(4,71)/(4,91)/(4,12)	southernMost	N	N
(4,71)/(4,92)	geographicName	N	Y
(4,71)/(4,92)/(4,13)	geographicKeywordName	Y	N
(4,71)/(4,92)/(4,14)	georgraphicKeywordType	Y	N
(4,93)	timePeriod	N	Y
(4,93)/(4,15)	TimePeriodStrutured	N	N
(4,93)/(4,16)	TimePeriodTextual	N	N
(4,98)	crossReference	N	Y
(4,98)/(4,50)	CrossReferenceTitle	Y	N
(4,98)/(4,17)	CrossReferenceLinkage	Y	N
(4,98)/(4,18)	CrossReferenceType	Y	N
(4,23)	originalControlIdentifier	Y	N
(4,59)	supplementalInformation	Y	N

Annex E (informative)

GILS Core Elements

GILS Locator Records consist of a number of GILS Core Elements that contain information to identify and describe Federal information resources. The term "mandatory" as used in this Profile applies to administration of the subset of GILS Locator Records that have been identified by the record source as participating in the GILS Core. GILS servers are not required to distinguish "mandatory" from other elements.

TITLE (Mandatory, Not Repeatable): This element conveys the most significant aspects of the referenced resource and is intended for initial presentation to users independently of other elements. It should provide sufficient information to allow users to make an initial decision on likely relevance. It should convey the most significant information available, including the general topic area, as well as a specific reference to the subject.

CONTROL IDENTIFIER (Mandatory, Not Repeatable): This element is defined by the information provider and is used to distinguish this locator record from all other GILS Core locator records. The control identifier should be distinguished with the record source agency acronym as provided in the U.S. Government Manual.

ABSTRACT (Mandatory, Not Repeatable): This element presents a narrative description of the information resource. This narrative should provide enough general information to allow the user to determine if the information resource has sufficient potential to warrant contacting the provider for further information. The abstract should not exceed 500 words in length.

PURPOSE (Mandatory, Not Repeatable): This element describes why the information resource is offered and identifies other programs, projects, and legislative actions wholly or partially responsible for the establishment or continued delivery of this information resource. It may include the origin and lineage of the information resource, and related information resources.

ORIGINATOR (Mandatory, Not Repeatable): This element identifies the information resource originator, named as in the U.S. Government Manual where applicable.

ACCESS CONSTRAINTS (Mandatory, Not Repeatable): This element in some cases may contain the value "None." It describes any constraints or legal prerequisites for accessing the information resource or its component products or services. This includes any access constraints applied to assure protection of privacy or intellectual property, and any other special restrictions or limitations on obtaining the information resource. Guidance on obtaining any users' manuals or other aids needed for the public to reasonably access the information resource must also be included here.

USE CONSTRAINTS (Mandatory, Not Repeatable): This element in some cases may contain the value "None." It describes any constraints or legal prerequisites for using the information resource or its component products or services. This includes any constraints applied to assure the protection of privacy or intellectual property and any other special restrictions or limitations on using the information resource.

AVAILABILITY (Mandatory, Repeatable): This element is a grouping of subelements that together describe

how the information resource is made available.

DISTRIBUTOR (Mandatory, Not Repeatable): This subelement consists of the following subordinate fields that provide information about the distributor:

DISTRIBUTOR NAME
DISTRIBUTOR ORGANIZATION
DISTRIBUTOR STREET ADDRESS
DISTRIBUTOR CITY
DISTRIBUTOR STATE
DISTRIBUTOR ZIP CODE
DISTRIBUTOR COUNTRY
DISTRIBUTOR NETWORK ADDRESS
DISTRIBUTOR HOURS OF SERVICE
DISTRIBUTOR TELEPHONE
DISTRIBUTOR FAX

RESOURCE DESCRIPTION (Optional, Not Repeatable): This subelement identifies the resource as it is known to the distributor.

ORDER PROCESS (Mandatory, Not Repeatable): This subelement provides information on how to obtain the information resource from this distributor, including any fees associated with acquisition of the product or use of the service, order options (e.g., available in print or digital forms, PC or Macintosh versions), order methods, payment alternatives, and delivery methods.

TECHNICAL PREREQUISITES (Optional, Not Repeatable): This subelement describes any technical prerequisites for use of the information resource as made available by this distributor.

AVAILABLE TIME PERIOD (Optional, Repeatable): This subelement provides the time period reference for the information resource as made available by this distributor, in one of two forms:

TIME PERIOD -- STRUCTURED: Time described using the USMARC prescribed structure

TIME PERIOD -- TEXTUAL: Time described textually.

AVAILABLE LINKAGE (Optional, Not Repeatable): This subelement provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information.

AVAILABLE LINKAGE TYPE (Optional, Not Repeatable): This subelement occurs if there is an Available Linkage described. It provides the data content type (i.e., MIME) for the referenced URI.

POINT OF CONTACT FOR FURTHER INFORMATION (Mandatory, Not Repeatable): This element identifies an organization, and a person where appropriate, serving as the point of contact plus methods that may be used to make contact. This element consists of the following subelement

CONTACT NAME
CONTACT ORGANIZATION
CONTACT STREET ADDRESS
CONTACT CITY
CONTACT STATE
CONTACT ZIP CODE
CONTACT COUNTRY
CONTACT NETWORK ADDRESS
CONTACT HOURS OF SERVICE
CONTACT TELEPHONE
CONTACT FAX.

RECORD SOURCE (Mandatory, Not Repeatable): This element identifies the organization, as named in the U.S. Government Manual, that created or last modified this locator record.

DATE OF LAST MODIFICATION (Mandatory, Not Repeatable): This element identifies the latest date on which this locator record was created or modified.

AGENCY PROGRAM (*, Not Repeatable): This element identifies the major agency program or mission supported by the system and should include a citation for any specific legislative authorities associated with this information resource. *This element is mandatory if the resource referenced by this GILS Core locator record is a Federal information system.

SOURCES OF DATA (*, Not Repeatable): This element identifies the primary sources or providers of data to the system, whether within or outside the agency. *This element is mandatory if the resource referenced by this IS Core locator record is a Federal information system.

CONTROLLED VOCABULARY (Optional, Repeatable): This element is a grouping of subelements that together provide any controlled vocabulary used to describe the resource and the source of that controlled vocabulary:

INDEX TERMS -- CONTROLLED (Optional, Not Repeatable): This subelement is a grouping of descriptive terms drawn from a controlled vocabulary source to aid users in locating entries of potential interest. Each term is provided in the subordinate repeating field
CONTROLLED TERM.

THESAURUS (Optional, Not Repeatable): This subelement provides the reference to a formally registered thesaurus or similar authoritative source of the controlled index terms. Notes on how to obtain electronic access to or copies of the referenced source should be provided, possibly through a Cross Reference to another locator record that more fully describes the standard and its potential application to locating GILS information.

LOCAL SUBJECT INDEX (Optional, Not Repeatable): This element is a grouping of descriptive terms to aid users in locating resources of potential interest, but the terms are not drawn from a formally registered controlled vocabulary source. Each term is provided in the repeating subelement:

LOCAL SUBJECT TERM

METHODOLOGY (Optional, Not Repeatable): This element identifies any specialized tools, techniques, or methodology used to produce this information resource. The validity, degree of reliability, and any known possibility of errors should also be described.

SPATIAL REFERENCE (Optional, Not Repeatable): This element is a grouping of subelements that together provide the geographic reference for the information resource. Geographic names and coordinates can be used to define the bounds of coverage. Although described here informally, the spatial object constructs should be as defined in FIPS 173, "Spatial Data Transfer Standard."

BOUNDING RECTANGLE (Optional, Not Repeatable): This subelement provides the limits of coverage expressed by latitude and longitude values in the order:

WESTERN-MOST
EASTERN-MOST
NORTHERN-MOST
SOUTHERN-MOST.

GEOGRAPHIC NAME (Optional, Repeatable): This subelement identifies significant areas and/or places within the coverage through two associated constructs:

GEOGRAPHIC KEYWORD NAME
GEOGRAPHIC KEYWORD TYPE.

TIME PERIOD OF CONTENT (Optional, Repeatable): This element provides time frames associated with the information resource, in one of two forms:

TIME PERIOD -- STRUCTURED: Time described using the USMARC prescribed structure

TIME PERIOD -- TEXTUAL: Time not described in the USMARC prescribed structure.

CROSS REFERENCE (Optional, Repeatable): This element is a grouping of subelements that together identify another locator record likely to be of interest:

CROSS REFERENCE TITLE (Mandatory, Not Repeatable): This subelement provides a human readable textual description of the cross reference.

CROSS REFERENCE LINKAGE (Mandatory, Not Repeatable): This subelement provides the machine readable information needed to perform the access (i.e., URI).

CROSS REFERENCE TYPE (Mandatory, Not Repeatable): This subelement occurs if there is a CROSS REFERENCE LINKAGE AND provides the data content type (i.e., MIME) for the referenced URI.

ORIGINAL CONTROL IDENTIFIER (Optional, Repeatable): This element is used by the record source agency to refer to another GILS locator record from which this locator record was derived.

SUPPLEMENTAL INFORMATION (Optional, Not Repeatable): Through this element, agencies may associate other descriptive information with the GILS Core locator record.

Attachment M

PROPOSED FEDERAL INFORMATION PROCESS STANDARD (FIPS) FOR APPLICATION PROFILE FOR THE GOVERNMENT INFORMATION LOCATOR SERVICE (GILS)

purposes mostly by graduate students seeking advance degrees. *Application Accepted by Commissioner of Customs:* May 31, 1994.

Docket Number: 94-077. *Applicant:* The Ohio State University, Aeronautical & Astronautical Research Laboratory, 2300 West Case Road, Columbus, OH 43235. *Instrument:* High Power, High Pressure Arc Discharge Plasma Source System. *Manufacturer:* Institute of Problems of Electrophysics, CIS. *Intended Use:* The instrument will be used for research that involves the determination of arc discharge mode, energy and power transfer efficiencies, electrode and insulator damage and wear. The experiments will be conducted for various gases over a range of initial pressures and currents and will involve electrical, pressure transducer and optical measurements. In addition, the instrument will contribute to educational objectives by providing the basis for student thesis research and design projects, and by challenging students with problems related to their coursework. *Application Accepted by Commissioner of Customs:* June 3, 1994.

Docket Number: 94-078. *Applicant:* University of Pittsburgh, Chemistry Department, 350 Thackeray Hall, Pittsburgh, PA 15260. *Instrument:* Mass Spectrometer, Model VG AutoSpec. *Manufacturer:* Fisons Instruments, United Kingdom. *Intended Use:* The instrument will be used to produce mass spectra of a wide variety of unique synthetic products and intermediate compounds (natural products, organic substrates for enzyme interaction, vitamin derivatives, and novel synthetic products) in the support of various research programs. The instrument will also be used to investigate the mechanisms of bombardment induced gas-phase ion chemistry and will involve the designing of additives to a FAB matrix that will induce specific chemical reactions within the mass spectrometer. In addition, the instrument will be used for educational purposes in the course Chemistry 2700 Graduate Research. *Application Accepted by Commissioner of Customs:* June 3, 1994.

Docket Number: 94-079. *Applicant:* Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439. *Instrument:* EPR Spectrometer, Model ESP300-E-10-12. *Manufacturer:* Bruker Instruments, Germany. *Intended Use:* The instrument will be used to study photo-induced charge separation in natural photosynthetic and model photosynthetic systems. The experiments to be conducted will be both conventional electron paramagnetic resonance experiments on stable radicals and time-resolved electron paramagnetic resonance. *Application Accepted by Commissioner of Customs:* June 3, 1994.

Docket Number: 94-080. *Applicant:* University of California, Los Alamos National Laboratory, P.O. Box 990, Los Alamos, NM 87545. *Instrument:* Electron Microscope, Model JEM 2010. *Manufacturer:* JEOL Ltd., Japan. *Intended Use:* The instrument will be used for the study of liquid crystal polymers and thermosets. Experiments will be performed on a variety of liquid crystalline materials to investigate the effect of different preparation schemes on the structure and the resulting properties. *Application Accepted by Commissioner of Customs:* June 7, 1994.

Pamela Woods,

Acting Director, Statutory Import Programs Staff.

[FR Doc. 94-16209 Filed 7-1-94; 8:45 am]

BILLING CODE 3510-DS-P

National Institute of Standards and Technology

[Docket No. 940680-4180]

RIN No. 0693-AB29

Proposed Federal Information Processing Standard (FIPS) for Application Profile for the Government Information Locator Service (GILS)

AGENCY: National Institute of Standards and Technology (NIST), Commerce.

ACTION: Notice; request for comments.

SUMMARY: This proposed Federal Information Processing Standard describes an application profile for the Government Information Locator Service (GILS). This application profile is based primarily on the American National Standard for Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (ANSI/NISO Z39.50-1992), developed by the National Information Standards Organization. The Government Information Locator Service (GILS) is a decentralized collection of servers and associated information services that will be used by the public either directly or through intermediaries to find public information throughout the Federal government.

Prior to the submission of this proposed FIPS to the Secretary of Commerce for review and approval, it is essential to assure that consideration is given to the needs and views of federal organizations, vendors, the public, and State and local governments. The purpose of this notice is to solicit such views.

The proposed FIPS contains two sections: (1) an announcement section, which provides information concerning the applicability, implementation, and maintenance of the standard; and (2) a specifications section which deals with the technical requirements of the standard.

Only the announcement section of the standard is provided in this notice. Interested parties may obtain copies of the technical specifications for this proposed FIPS for Application Profile for the Government Information Locator Service (GILS) from Standards Processing Coordinator (ADP), Computer Systems Laboratory, National Institute of Standards and Technology, Technology Building, Room B-64, Gaithersburg, MD 20899, telephone (301) 975-2816.

DATES: Comments on this proposed FIPS must be received on or before October 3, 1994.

ADDRESSES: Written comments concerning the proposed FIPS should be sent to: Director, Computer Systems Laboratory, ATTN: Proposed FIPS for GILS, Technology Building, Room B154, National Institute of Standards and Technology, Gaithersburg, MD 20899.

Written comments received in response to this notice will be made part of the public record and will be made available for inspection and copying in the Central Reference and Records Inspection Facility, room 6020, Herbert C. Hoover Building, 14th Street between Pennsylvania and Constitution Avenues, NW, Washington, DC 20230.

FOR FURTHER INFORMATION CONTACT: Mr. Eliot Christian, U.S. Geological Survey, 802 National Center, Reston, VA 22092, telephone (703) 648-7243, fax (703) 648-7069, E-mail: echristi@usgs.gov.

Dated: June 28, 1994.

Samuel Kramer,

Associate Director.

Proposed Federal Information Processing Standards Publication_____

(date)

Announcing the Standard for Application Profile for the Government Information Locator Service (GILS)

Federal Information Processing Standards Publications (FIPS PUBS) are issued by the National Institute of Standards and Technology (NIST) after approval by the Secretary of Commerce pursuant to Section 111(d) of the Federal Property and Administrative Services Act of 1949 as amended by the

Computer Security Act of 1987, Public Law 100–235.

1. Name of Standard. Application Profile for the Government Information Locator Service (GILS).

2. Category of Standard. Software Standard, Information Interchange.

3. Explanation. This standard describes an application profile for the Government Information Locator Service (GILS). This application profile is based primarily on the American National Standard for Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (ANSI/NISO Z39.50–1992), developed by the National Information Standards Organization (NISO). The Government Information Locator Service (GILS) is a decentralized collection of servers and associated information services that will be used by the public either directly or through intermediaries to find public information throughout the Federal government.

This GILS Profile specifies the use of ANSI/NISO Z39.50–1992 in information service applications and provides specifications for the overall GILS application, including the GILS Core and other aspect of a GILS server operating in the Internet environment. This GILS profile will enable GILS client systems to interconnect and to interoperate with any GILS server. This profile addresses intersystem interactions and information interchange for the GILS, but does not specify user interface requirements, the internal structure of databases that contain GILS Locator Records, or search engine functionality.

GILS servers will support search and retrieval by accepting a search query and returning a result set or diagnostic messages. GILS servers may also support browsing by accepting a well-known search query and returning a list of Locator Records in brief display format.

Some of the information resources pointed to by GILS Locator Records, as well as the GILS server itself, may be available electronically through other communications protocols including the common Internet protocols that facilitate electronic information transfer such as remote login (Telnet), File Transfer Protocol (FTP), and electronic mail. The use of SMTP and MIME protocols or other communications paths is outside the scope of the GILS Profile.

The GILS Profile was developed by a group of industry and government experts in ANSI/NISO Z39.50–1992 implementations, system implementations, and the organization of information. The specifications included in the GILS Profile reflect the consensus of this group based on its work and input from a range of stakeholders.

4. Approving Authority. Secretary of Commerce.

5. Maintenance Agency. U.S. Department of the Interior, United States Geological Survey (USGS).

Questions concerning this standard are to be addressed to the Maintenance Agency: GILS Program, United States Geological Survey (USGS), 802 National Center, Reston, VA 22092. Users of this standard who need to be notified or changes that occur prior to the next publication of the standard should complete the Change Request Form provided in this publication and send it to: Standards Processing Coordinator (ADP), Computer Systems Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899. The NIST will issue Change Notices on an as-needed basis.

6. Related Documents.

a. Federal Information Resources Management Regulations (FIRMR) subpart 201–20.303, Standards and subpart 201–39–1002, Federal Standards.

b. Office of Management and Budget Bulletin 94–_____, Establishment of Government Information Locator Service.

c. American National Standard for Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection (ANSI/NISO Z39.50–1992).

d. A list of additional references for the Application Profile is contained in section 5, References, of the specifications.

7. Objectives. The objectives of the Application Profile for the GILS are to:

—Enable users to identify, locate, and access or acquire publicly available Federal information resources, including electronic information resources.

—Provide a uniform approach to providing information locator services to the public.

—Enable every agency to establish standards-based network-accessible locator records.

8. Applicability.

a. This standard is recommended for use by Federal agencies in the development and establishment of information locators, i.e., information resources that identify other information resources, describe the information available in those resources, and provide assistance in how to obtain the information.

b. This standard is required for use by Federal agencies in those information locators that are established and maintained as part of the Government Information Locator System (GILS) pursuant to the requirements of OMB Bulletin 94–_____ and other applicable, law, regulation, and policy.

c. The GILS Core requirements of this standard apply to those GILS locator records which:

—Describe information resources maintained by the Federal government;

—Comply with the defined GILS Core Elements;

—Are mutually accessible through interconnected electronic network facilities without charge to the direct user; and

—Are designated by the agency to be part of the Federal government GILS Core, pursuant to OMB Bulletin 94–_____.

9. Specifications. The Application Profile for the Government Information Locator System, (affixed).

10. Implementation. The implementation of this standard involves three areas of consideration: development and acquisition of GILS implementations; validation, and interpretations of the standard.

10.1 Development and Acquisition of GILS Implementations. This standard is effective _____ (6 months after approval by the Secretary of Commerce).

10.2 Validation. Validation of GILS implementations is not required at this time. Testing for conformance to this standard is at the discretion of the agency. Agencies may select the tests to be administered and the testing organizations that administer the tests.

10.3 Interpretation of this standard. Resolution of questions regarding this standard will be provided by NIST. Questions concerning the content and specifications should be addressed to: Director, Computer Systems Laboratory, Attn: FIPS for GILS Interpretation, National Institute of Standards and Technology, Gaithersburg, MD 20899, Telephone: (301) 975–2833.

11. Waivers. Under certain exceptional circumstances, the heads of Federal departments and agencies may approve waivers to Federal Information Processing Standards (FIPS). The head of such agency may redelegate such authority only to a senior official designated pursuant to Section 3506(b) of Title 44, U.S. Code. Waivers shall be granted only when:

a. Compliance with a standard would adversely affect the accomplishment of the mission of an operator of a Federal computer system, or

b. Cause a major adverse financial impact on the operator which is not offset by governmentwide savings.

Agency heads may act upon a written waiver request containing the information detailed above. Agency heads may also act without a written waiver request when they determine that conditions for meeting the standard cannot be met. Agency heads may approve waivers only by a written decision which explains the basis on which the agency head made the required finding(s). A copy of each such decision, with procurement sensitive or classified portions clearly identified, shall be sent to: National Institute of Standards and Technology; Attn: FIPS Waiver Decisions, Technology Building, room B-154; Gaithersburg, MD 20899.

In addition, notice of each waiver granted and each delegation of authority to approve waivers shall be sent promptly to the Committee on Government Operations of the House of Representatives and the Committee on Governmental Affairs of the Senate and shall be published promptly in the Federal Register.

When the determination on a waiver applies to the procurement of equipment and/or services, a notice of the waiver determination must be published in the Commerce Business Daily as part of the notice of solicitation for offers of an acquisition or, if the waiver determination is made after that notice is published, by amendment to such notice.

A copy of the waiver, any supporting documents, the document approving the waiver and any supporting and accompanying documents, with such deletions as the agency is authorized and decides to make under 5 U.S.C. 552(b), shall be part of the procurement documentation and retained by the agency.

12. Where to Obtain Copies. Copies of this publication are for sale by the National Technical Information Service, U.S. Department of Commerce, Springfield, VA 22161. (Sale of the included specifications document is by arrangement with the United States Geological Survey (USGS).) When ordering, refer to Federal Information Processing Standards Publication _____ (FIPSPUB _____), and title. Payment may be made by check, money order, or deposit account.

[FR Doc. 94-16205 Filed 7-1-94; 8:45 am]

BILLING CODE 3510-CN-M

[Docket No. 940670-4170]

RIN 0693-AB26

Proposed Revision of Federal Information Processing Standard (FIPS) 125-1, MUMPS (Massachusetts General Hospital Utility Multi-Programming System)

AGENCY: National Institute of Standards and Technology (NIST), Commerce.

ACTION: Notice; Request for comments.

SUMMARY: This proposed revision of Federal Information Processing Standard (FIPS) 125-1, MUMPS (Massachusetts General Hospital Utility Multi-Programming System), will adopt the revised voluntary industry specifications, ANSI/MDC X11.1-199X. The American National Standard for M (also known as MUMPS, [MASSACHUSETTS GENERAL HOSPITAL UTILITY MULTI-PROGRAMMING SYSTEM]) specifies the form and establishes the interpretation of programs written in the M programming language.

Prior to the submission of this proposed revision to the Secretary of Commerce for review and approval, it is essential to assure that consideration is given to the needs and views of manufacturers, the public, and state and local governments. The purpose of this notice is to solicit such views.

This proposed FIPS contains two sections: (1) An announcement section, which provides information concerning the applicability, implementation, and maintenance of the standard; and (2) a specifications section which deals with the technical requirements of the standard. Only the announcement section of the standard is provided in this notice. Interested parties may obtain copies of the technical specifications (ANSI/MDC X11.1-199X) from the MUMPS Development Committee (MDC) Secretariat, 1738 Elton Road, Suite 205, Silver Spring, MD 20903, (301) 431-4070, FAX (301) 431-0017.

DATES: Comments on this proposed revision must be received on or before October 3, 1994.

ADDRESSES: Written comments concerning the proposed revision should be sent to: Director, Computer Systems Laboratory, ATTN: Proposed FIPS 125-2, M, Technology Building, Room B-154, National Institute of Standards and Technology, Gaithersburg, MD 20899.

Written comments received in response to this notice will be made part of the public record and will be made available for inspection and copying in the Central Reference and Records Inspection Facility, Room 6020, Herbert C. Hoover Building, 14th Street between Pennsylvania and Constitution Avenues, NW., Washington, DC 20230.

FOR FURTHER INFORMATION CONTACT: Dr. William H. Dashiell, National Institute of Standards and Technology, Gaithersburg, MD 20899, (301) 975-2490.

Dated: June 27, 1994.

Samuel Kramer,

Associate Director.

Proposed Federal Information Processing Standards Publication 125-2 (Supersedes FIPS PUB 125-1—1993 June 10)

(date)

Announcing the Standard for M (Also Known as MUMPS [MASSACHUSETTS GENERAL HOSPITAL UTILITY MULTI-PROGRAMMING SYSTEM])

Federal Information Processing Standards Publications (FIPS PUBS) are issued by the National Institute of Standards and Technology (NIST) after approval by the Secretary of Commerce pursuant to Section 111(d) of the Federal Property and Administrative Services Act of 1949, as amended by the Computer Security Act of 1987, Public Law 100-235.

1. Name of Standard. M (also known as MUMPS [MASSACHUSETTS GENERAL HOSPITAL UTILITY MULTI-PROGRAMMING SYSTEM]) (FIPS PUB 125-2).

2. Category of Standard. Software Standard, Programming Language.

3. Explanation. This publication announces the adoption of American National Standard for M, ANSI/MDC X11.1-199X, as a Federal Information Processing Standard (FIPS). The American National Standard for M, ANSI/MDC X11.1-199X, specifies the form and establishes the interpretation of programs written in the M programming language. The purpose of the standard is to promote portability of M programs for use on a variety of data processing systems. The standard is for use by implementors as the reference authority in developing compilers, interpreters, or other forms of high level language processors; and by other computer professionals who need to know the precise syntactic and semantic rules adopted by ANSI. This publication is a revision of FIPS PUB 125-1 and supersedes that document in its entirety.

4. Approving Authority. Secretary of Commerce.

5. Maintenance Agency. U.S. Department of Commerce, National

Attachment N

RESPONSE TO STAKEHOLDER ON SUITABILITY OF Z39.50 FOR GILS

Dear _____:

As I mentioned, I passed your messages on to two of the people who worked on the GILS Profile Development project, Denis Lynch and Ralph Levan. Both of them have been involved extensively with Z39.50 implementations as well as in the development of the standard. I've compiled their comments and responses.

GILS is a development project whose purpose is to improve citizens' access to important Federal information sources *THIS YEAR*. It has been established that online access, particularly via the Internet, is required, which means that a protocol must be used. Since the government would like an open-systems, standards-based solution for GILS, there are two choices:

1) Develop a new protocol

2) Use an existing protocol.

Clearly (1) won't work; it would take longer than "this year" just to design the protocol, never mind getting it implemented in software from multiple vendors.

So that leaves us with (2). Now we're reduced to shopping, rather than designing. The choices I'm aware of are:

1) Gopher
2) WAIS
3) HTTP
4) SFQL
5) Z39.50

1) Gopher is very widely available, with many independent implementations. But its notions of searching and information transmission just aren't adequate for GILS.

2) WAIS has only been implemented once. Various minor versions of the user interface exist, but they are all based on one code body. This is unlikely to change, since the WAIS protocol is a unique -- and essentially undocumented -- variation of a very limited subset of an old version of Z39.50 (i.e., Z39.50-1998). Current versions of Z39.50 cover all of WAIS needs, so the variation is no longer necessary.

3) HTTP handles searching only incidently, and only deals with one data type - HTML. (GILS records could be easily encoded in HTML documents, so HTML isn't the big problem -- searching is).

4) SFQL handles searching in the same rather restricted way as SQL. It doesn't address data transmission at all. It is used only in some very exclusive communities.

5) Z39.50 is complex, and has some of the shades of green that come from association with the OSI dinosaur. But it is used (or about to be used) by a very wide audience, with software from quite a few unrelated providers. (An editorial aside: Z39.50 is the work of wide community of information providers and users. Brewster Kahle has intermittently been a member of the community, but only one member. "Following the development of WAIS" is orthogonal to following the development of Z39.50.)

Those of us who have studied the tradeoffs believe that Z39.50 is the right choice for GILS. If there are other chioces it might not be too late, but none have surfaced.

Now about the general "goodness" of Z39.50 and its suitability for the GILS application:

The most important aspect of Z39.50 is that it recognizes that information access comes in two phases -- selection and retrieval -- and provides a common language that covers both phases. The phases are quite separate, and each phase allows for considerable variation. At the simplest level, the common language is:

- Establish connection
- Select items
- Transfer items

The "search" service is used to communicate the selection criteria. The standard specifies one widely-used language ("RPN Query") for specifying the criteria, but others can be added as needed. An RPN Query specifies a set of constraints to be met by interesting items. The "RPN" has to do with how the search engine is to understand the meaning of the query **NOT** how the search engine should execute the query!

The individual constraints (operands) in an RPN query have two parts:

- some information
- how that information is to be used.

The information can be a simple text string, but it can also be any data structure you can define -- no limits at all. The "how it is to be used" commonly consists of a database access point (sometimes thought of as an index, but that's a specific implementation approach) and how the access point and client-supplied information are to relate (Begins With, Is Not Equal To the Author Named, Is Relevant To). The protocol doesn't distinguish between the access point and the relation specifiers -- they are all lumped into "how should the information you sent be applied to finding interesting items?". New relationship vocabularies can be -- and have been -- developed as needed.

A current project is developing a Z39.50 interface to Conquest's semantic-net based dictionary, so there isn't a protocol issue involved with the set of categories to be used in GILS records. There is a separate issue of the "content" of the records (will they use a controlled vocabulary, and if so how will the vocabulary develop over time). This is a significant concern, but not related to the choice of protocol.

On the information transmission side: One of the things that Z39.50 does better than anything else we're aware of is to provide the client and server with a mechanism to find out what "information formats" are commonly understood. All the formats you've grown accustomed to (raw text, graphics, HTML, etc.) will be handled nicely by many systems -- certainly the clients that some of us are building! HTML links (really URLs or URNs) will be handled likely by almost everybody.

URNs are an IETF development, currently underway. One might think of them as intending to be "dynamic hyperlinks." Z39.50 will carry them, however they are eventually defined. In any case, URNs will be able to refer to items that are accessible via Z39.50 as well as ones that aren't. That's just like in the Web -- links can point to things that are accessed by lots of different mechanisms.

About complexity: V3D9, or whatever you've seen, is not easy sledding. It describes a pretty big protocol with lots of options. Few of those options are needed to use GILS. (The equivalent of Rose's "The Open Book" could be a big seller.) It really isn't hard if you take things one at a time. In any case, the "readability of a standard" definition is probably not a good measure of the standard's applicability for a specific purpose! The number of implementors is probably a better one, and there may be more distinct implementations of Z39.50 than TCP.

About EDI: EDI is an SGML application that describes a data format for storing purchase order-like things. It isn't a protocol. Z39.50 will have mechanisms for submitting records, and the format can be whatever the client and server mutually agree is acceptable. For GILS records it's easy to imagine an SGML DTD, perhaps MARC records, and other more information-dense formats. Z39.50 *is* a way of automatically requesting records, so that part is taken care of from the start. On the other hand, if somebody wanted to implement it, there's no reason that an EDI form couldn't be defined as an acceptable query format!

Summarizing, Z39.50 is a protocol that allows information consumers to communicate with information providers. The protocol allows very simple or very smart software to exist at either end. The protocol is totally neutral about the organization and meaning of the data records.

Now a GILS server, and the overall GILS structure, is another matter, separate from the utility and function of Z39.50. Is the current non-definition of things like category terms a problem? It is our understanding that the Federal government would like to be able to offer a open systems, standards-based mechanism for people to identify and locate publicly available government information, and that it is it's more important to get *some* data up for people to use than to try to establish the bureaucracy and procedures it would take to control that sort of thing.

Should GILS servers be able to do amazingly clever information retrieval? You bet. We expect that better retrieval performance will be one of the major distinguishers among commercial information providers! But that will be best solved in the marketplace, not in a requirements document!

In response to some of your specific points:

> 1) extend query-type to support queries on a knowledge representation
> 2) make hierarchical categories allow relational links and constraints

Z39.50 does have such extensibility, and the natural language folks that are working with the Z39.50 Implementors Group (ZIG) do not seem to have any problems with it. Some of your concerns here may be due to looking at WAIS (based on the subset of the 1988 version of the protocol) rather than at the current version of Z39.50 and operational implementations. As we suggested above, you can stick almost anything in Z39.50 queries.

> 3) be able to return html pointers and pages (should be easy and already seen
> in various places).

Z39.50 can do that now. But, as nice as HTML looks today, it is FAR from stable and one might be cautioned from making any 5-year plans based on it's continued existence. This stuff is just too ad-hoc and dynamic.

> 4) plan for extension to on-the fly linking of documents.

That's a document contents and client usage problem. In the specific case of GILS, we've defined how this can be done using URL's in well tagged fields.

> 5) the protocols need to be easy to implement, widely available. This probably
> means a small required core.

While "easy to implement" may be an attractive goal, it's not likely — especially when the functional requirements become more complex. Gopher is easy to implement, and it won't do what we want. But, the wide availability of independently developed Z39.50 clients and servers is a strong argument that Z39.50 objects are not that hard to implement. And eventually there will be plenty of Z39.50 API's to build on. There at least four now (available from Stanford, National Library of Canada, OCLC, and CNIDR).

> 6) the protocol should be applied recursively so that one protocol gets you
> everything.

We're working on that. That is what some people believe the Explain facility can become, but some of us don't agree with that. This problem is a little too hard to lump into an already complex standard. Maybe later when the solutions to these problems are a little better understood.

> This is crucial because a public interface will need to be top-down point and click because most people will
> be unable to formulate effective search keys.

> Naturally, this then links back to a taxonomic characterization of all government information resources.
> So, WAIS will need a serious top-down component to complement the bottom up approach that it has
> historically implemented.

The problem here is the government's unwillingness/inability to create such a taxonomy. Too many diverse organizations that cannot be coerced/motivated into cooperating. And, this is not WAIS.

> I forget whether Z39.50 has anything that can be coerced into a typed link. I think not, and this is a serious
> shortcoming not just for the distributed hypertext application, not to mention the distributed knowledge
> representation case.

This is a record data problem, which has nothing to do with Z39.50. In the case of GILS, its records have all the links we could think of, and others can be added in an ad-hoc manner.

> Z39.50v3 is 100 pages. It should have an appendix detailing what minimal conformance means.

That's what profiling is about, and the GILS profile provides just such minimal conformance information.

> I don't think we very near a convincing answer to the evolution issue at this point, and I suspect we will
> need some serious study to determine the answer.

While we agree with that, we also don't have years to spend coming up with the perfect solution. Z39.50 will evolve along with our mutual understanding of how people need/want to find information.

> 1960S CONCEPTS IN 1980S NETWORK CLOTHING:

Boolean Algebra has it's place in information retrieval. It does not lend itself well to Natural Language queries (nor does SQL for that matter), but Z39.50 queries allow for arbitrary structure in the terms; they need not be atomic. The Natural Language folks we're working with are also quite happy to be able to mix the specificity of term based booleans with the flexibility of Natural Language.

Z39.50 has evolved in recent years to meet the demands of information providers that have operational and production systems in place. The original focus of Z39.50 on searching and retrieving a particular type of data, namely bibliographic records, has been expanded dramatically, precisely because of people wanting to search and retrieve against other types of document databases. And it has been the extensibility of the standard that has allowed this evolution. While no tool is perfect, the implementation base of Z39.50 is growing rapidly, and we feel that the choice by the Federal government to use this protocol in no way constrains future developments of GILS. In part, because of the extensibility of the protocol, but more importantly, GILS servers and clients, the search engines, database structures, and advanced IR techniques that may be implemented are all separate from the what Z39.50, as a peer-to-peer communications protocol, addresses.

I hope that this information is helpful and responds to some of the basic concerns and questions you voiced. Let me know if there is anything else.

Cheers,
Bill Moen, GILS Profile Development Project Manager
School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244
wemoen@mailbox.syr.edu
(315) 443-4508/445-0015

Attachment O

USMARC PROPOSAL 94-9: CHANGES TO THE USMARC BIBLIOGRAPHIC FORMAT TO ACCOMMODATE ONLINE SYSTEMS AND SERVICES

PROPOSAL NO: 94-9

DATE: May 6, 1994
REVISED: July 20, 1994

NAME: Changes to the USMARC Bibliographic Format to Accommodate Online Systems and Services

SOURCE: Library of Congress; OCLC Internet Resources Project

SUMMARY: This paper proposes several enhancements to the USMARC Bibliographic Format to allow for the creation of USMARC records for online systems and services. Included are the following: addition of a code for online system and service in 008/26 (Type of computer file) for the USMARC computer file specifications; discussion of the use of a code in field 042 to identify records as part of the GILS project and 040 $e (Description conventions) to indicate application of the Guidelines for Describing Internet Resources; Addition of Community Information Format fields 270 (Primary Address), 301 (Hours, Etc.) (option to use tag 307), and 531 (Eligibility, Fees, Procedures Note) for use in USMARC bibliographic records; Addition of subfield $w for Record control number in field 856 (Electronic Location and Access) for linking from a record for an electronic data resource to the record for the online system and service; a discussion of the Uniform Resource Name (URN) and how it might apply to this type of record.

KEYWORDS: Field 008/26 (Computer files); Type of Computer File; Field 042; Authentication Code; Subfield $e (040 Bibliographic); Field 270 (Bibliographic/Community Information); Address; Field 301 (Bibliographic/Community Information); Hours, Etc.; Field 531 (Bibliographic/Community Information); Eligibility, Fees, Procedures Note; Subfield $w (856 Holdings/Bibliographic); Record Control Number; Uniform Resource Name; URN

RELATED: DP 49 (June 1991); DP 54 (Jan. 1992); 93-4 (Jan. 1993); DP 69 (June 1993); DP 78 (June 1994)

DATES	STATUS/COMMENTS
5/6/94	Forwarded to USMARC Advisory Group for discussion at the June 1994 MARBI meetings.
6/25/94	Results of USMARC Advisory Group discussion: Approved as amended. Amendments are as follows: **1. Field 270 (Primary address).** Change name to Address. Add subfield $z (Public note (R)) and subfield $r (Hours (R)) for hours contact is available. **2. Field 301/307 (Hours, etc.).** Use field tag 307. Do not define $c but leave time zone to be indicated informally in $a or $b. **3. Field 531 (Eligibility, Fees, Procedures Note).** Do not add 531. LC will initiate a future proposal to consider using field 506 for this data. **4. Field 856 $w (Record control number).** Add description to clarify that the linkage

is field-to-record, not record-to-record (as in linking entry fields). Change examples showing a repeatable $2; it should not be repeatable.
LC will initiate proposals as appropriate to consider the following in the future:
1. Align Community Information Format with changes in this proposal to field 270 and 301/307.
2. Consider adding subfield in field 856 for hours access method is available.
3. Possible field 506 changes (see above).
4. Consider punctuation of the description in examples where no ISBD punctuation is shown.
In addition, LC should consider coordinating subfield $2 (Access method) in field 856 with Uniform Resource Locators (URLs).

7/20/94 Results of final LC review:
Agreed with the MARBI decision.

Proposal No. 94-9: Changes to the USMARC Bibliographic Format to Accommodate Online Systems and Services

I. BACKGROUND

The USMARC Advisory Group has discussed several papers related to accommodating online information resources in USMARC. Discussion Paper 49 (Dictionary of Data Elements for Online Information Resources), discussed in June 1991, presented the data elements needed for online information resources and gave a tentative mapping to USMARC bibliographic fields. Participants agreed that USMARC should be expanded to accommodate description and access of machines as resources on the network as well as data files on the machines, and that further work on the data elements and USMARC mapping needed to be done. Discussion Paper No. 54 (Providing Access to Online Information Resources) introduced questions of scope and the use of fields in the USMARC Holdings Format and the new Community Information Format (provisionally approved during Midwinter 1992 ALA). It was agreed that electronic data resources (e.g. electronic text, software, data files, bibliographic databases, electronic graphics files) might be more amenable than online systems and services (e.g., FTP sites, Telnet sites, listservs, bulletin boards, campuswide information systems) to bibliographic description using AACR2 computer files cataloging rules and the USMARC bibliographic format as they now exist, and that more work needs to be done to accommodate online systems and services.

Proposal No. 93-4 (Changes to the USMARC Bibliographic Format (Computer Files) to Accommodate Online Information Resources) attempted to accommodate electronic data resources in the USMARC format using the computer files specifications. It proposed adding codes and changing some definitions in 008/26 (Type of computer file) to better identify these items; broadening the use of field 256 (File Characteristics) to include more specific descriptors; making field 516 (Type of File or Data Note) obsolete; and adding a new field 856 to the Holdings/Bibliographic formats for electronic location and access information. It did not specifically cover online systems and services, although field 856 was designed to accommodate them. The MARBI discussion resulted in the approval of changing the 008/26 with some amendments, and approval of the addition of field 856 (with some amendments) as a provisional field, pending experimentation on its use. (The 256 and 516 changes have been dropped because of the effect on the cataloging rules and the decision of a task force of the Committee on Cataloging: Description and Access (CC:DA) not to change the rules.)

The USMARC Advisory Group discussed Discussion Paper No. 69 (Accommodating Online Systems and Services in USMARC) in June 1993. The consensus of the group was that the Library of Congress should prepare a proposal to allow for the creation of MARC records for online systems and services. Following summarizes the discussion:

- The bibliographic format should be used, since there is an interrelationship between online systems/services and bibliographic records. Also, the Community Information Format is not that well defined, and records in that format may not be maintained by libraries.
- The records need to be identified, since some institutions will include them in the same database with bibliographic materials, and others may wish to include them in a separate database or directory and use like an enhanced Gopher. Identification should probably be in 008
- Some Community Information Format fields should be defined for use in the USMARC Bibliographic Format for data elements that are not currently accommodated.

- Linking techniques need to be discussed further. Information that applies universally to the online system or service should be distinguished from local information, perhaps in field 856.

To fully accommodate online systems and services, it is necessary to provide information on location and access for non-Internet resources, i.e., those accessible through dial-up. Discussion Paper 78 (Location and Access Information for non-Internet resources in USMARC Records) discusses this issue.

2. Government Information Locator Service

The Government Information Locator Service (GILS) has been established to help the public locate and access information throughout the U.S. government. This is a locator system to identify databases and services that provide government information, and will also includes electronic data resources and in some cases printed information. Federal agencies are organizing GILS as a component of the National Information Infrastructure (NII). It is intended to make government information available electronically by identifying, describing and providing access information to locations where information resides. Federal agencies will be responsible for participation in GILS by providing locator records; it is left up to the agency to what level they wish to describe their resources, whether at a high level only describing large systems, or at a lower level describing many types of their resources. Many GILS records would describe online systems and services; thus it is a subset of those types of online information resources.

GILS will use the information search and retrieval standard known in the United States as ANSI/NISO Z39.50 (known internationally as ISO 10162/10163). Locator records are to be available in three specified formats, one of which is USMARC. Consequently, an effort has been underway to map GILS data elements to the USMARC Format for Bibliographic Data. Data elements have been defined and appropriate fields indicated. In most cases no new fields were used in the mapping although some USMARC definitions have been expanded. GILS would operate through decentralized servers using Z39.50 to navigate. Each agency would build records for their own resources and may bring in records from elsewhere. It was intentionally designed so that agencies could take some other agency's record and enhance or change it so that it accommodates the local agency's needs; cross references between the record in different systems is possible. Implementation issues have not fully been resolved.

The changes suggested in this proposal would better accommodate the GILS data into USMARC than the current mapping does. They would allow for clearer identification of the record and more structured recording of contact information.

3. Identifying USMARC records for online systems and services

Discussion Paper No. 69 explored the use of the bibliographic or community information formats for the creation of USMARC records for online systems and services. The discussion revealed that participants felt that, in order to incorporate these records into library catalogs along with other electronic resources, they should be in the computer files format. Thus, records for "electronic data resources" (e.g., text, software, bibliographic databases) would reside in the same system or database as records for "online systems and services" (e.g., campus-wide information systems, online services, bulletin boards, etc.). An additional advantage to this approach is that then the record for the electronic data resource could link to the record for the online system so that electronic location information would not need to be repeated

and thus maintained in each individual resource record. Field 856 (Electronic Location and Access in the record for the online service only would need to be kept up-to-date.

Field 008/26 (Type of computer file) in the computer file specifications of the bibliographic format could be used to identify a record for an online system and service as such. This information would alert the computer in cases where online systems records have special processing or need to be extracted for special purposes. It also would alert the user to the fact that the record for an online system may have special characteristics since it is not a bibliographic entity (e.g., use of community information fields, AACR2 cataloging rules not applicable, etc.). Code "i" could be defined for "online system or service". Code "e" would need to be slightly reworded to distinguish between "bibliographic data" and "online system or service". (See Attachment A)

A message distributed on the USMARC discussion list explored questions about the kind of information needed about GILS records to identify them and to indicate that the record may not use full content designation or cataloging rules. These questions may also apply to other online systems and services records. The records could include information to further show their status as less than full MARC/AACR2 cataloging records:

- Leader/18 (Descriptive cataloging form) could be set to # (Non-ISBD) if International Standard Bibliographic Description is not followed; this could be used for something like a GILS record where no attempt is made to provide standard ISBD punctuation. In other cases a GILS (or other online system record) could be set to "a" (AACR2) if cataloging were consistent with AACR2.
- 040$e (Description conventions) could contain information about descriptive rules used in creating the record. A code could be defined for the new Guidelines for Cataloging Internet Resources, written as part of the OCLC Internet Resources Project and recently approved (with a few changes) by the ALCTS Committee on Cataloging: Description and Access (CC:DA). It will probably be necessary to reconsider the guidelines in terms of creating records for the online systems and services, since it was written with "electronic data resources" in mind. OCLC plans to publish the guidelines in the future.
- Applicable codes could be defined in field 042 (Authentication code) to indicate that the record was derived from a specific project or agency (e.g. GILS). The definition of the code could include information on whether or not headings are verified against an authority file.

4. Defining Community Information Format fields in the USMARC bibliographic format

Discussion Paper No. 69 included a mapping for data elements needed to create records for online systems and services to USMARC fields. The paper identified three fields from the community information format that could be used for description of online systems and services. These are:

Hours of Service	CIF 301 $a (Hours, etc.)
Telephone	CIF 270 $k (Telephone number) or $j (Specialized telephone number)
Fax	CIF 270 $l (Fax)
Cost for Use	CIF 531 $b (Fees)

The GILS profile includes data elements for point of contact, breaking it into subelements as follows.

The Community Information Format includes separate subfields for most of these data elements. Following each is the USMARC Community Information Format subfield/field that is appropriate:

Point of Contact	Community Information Format field
Contact Name	270$p
Contact Organization	270$p
Contact Street Address	270$a
Contact City	270$b
Contact State	270$c
Contact Zip Code	270$e
Contact Country	270$d
Contact Network Address	270$m
Contact Hours of Service	301$a
Contact Telephone	270$k
Contact FAX	270$l

Although these data elements are applicable to GILS, they would also be appropriate for other types of online systems and services records.

Note that field 301 (Hours, etc.) was defined as Physical Description for Visual Materials in the USMARC Bibliographic format and was made obsolete in 1983. It could be redefined as 301 to be consistent with the Community Information Format, or a new tag, field 307, could be assigned (and then 301 changed in CIF).

An alternative to defining the Community Information Format fields would be to use 856$m (Electronic Location and Access--Contact) for the data. However, it could not be parsed into separate subelements but all included in this subfield. It is unlikely that any special processing would be done from the data, but it is possible that a display might be desired. If this approach were followed, there would be no standard way to include the data, and there would be a lot of data in the subfield. The GILS to MARC mapping used 856$m for contact information for electronic resources and 535 (Location of Originals/Duplicates Note), which has subfields for addresses, for non-electronic resources, because the Community Information Format fields were not yet available in the bibliographic format. If these fields were defined, GILS could use them regardless of the form of the resource (electronic or non-electronic).

5. Linkage of electronic data resources records to online systems and services records.

If USMARC records are created for online systems and services and incorporated into the library catalog, the record could include the electronic location and access information for that service. Any electronic data resource (e.g. text files, software, etc.) that can be accessed at the location could be linked to the record for the system or service, rather than the data included in each individual record. It is possible that in the future the location information could be kept up-to-date in only the system or service record when the tools are available to do so (i.e., the ability to resolve names and locations through directory services). In order to link the two types of USMARC records, a subfield $w for Record control number is needed in field 856. It would contain the record control number for the online system and service record, so that one could then find the location and access information for the particular resource. Subfield $w is also available in the 76X-78X Linking Entry fields as System control number for linkage.

6. Uniform Resource Name (URN) and USMARC

The Internet Engineering Task Force (IETF) is developing a family of standards called Uniform Resource Identification (URI) to identify, describe, locate, and control networked information objects on the Internet. The Uniform Resource Locator (URL) is the address of an object, containing enough information to identify a protocol to retrieve the object. Elements of the draft URL standard are contained in separate subfields of field 856 in USMARC; in the URL the elements are strung together with separators between them. If an institution wishes to use the URL as it has been established, the new URL subfield (856$u) is used. An institution may wish to record only the URL, rather than use the separate subfields, record both parsed elements and the URL, or record only the parsed subfields. Recording the elements in separate subfields may be useful to create a display or to verify the separate data elements even if the URL is also used.

The Uniform Resource Name (URN) is intended to be a persistent, location independent identifier for an object, providing a unique element to identify it. The URN will "provide a globally unique, persistent identifier used both for recognition and often for access to characteristics of or access to the resource". It may identify "intellectual content or a particular presentation of intellectual content", depending upon how the assignment agency uses it. A resource identified by a URN may reside at many locations under any number of filenames and may move any number of times during its lifetime. The URL identifies the location for an instance of a resource identified by the URN. The URN is still under development and not all issues have been resolved. When it is finalized, it will provide bibliographic control to uniquely identify a resource. It will have an impact on the decision as to when to consider a resource a new edition and thus create a separate record, for it attempts to identify whether two items are the same or different in intellectual content.

For purposes of this paper, it is important to be aware of the development of the URN, but at this point it has not been sufficiently defined to attempt to accommodate it in USMARC. The group developing the standard continues to discuss when a new URN is assigned and when it is considered the same as another. The major factor in making this decision that the IETF is using is how the assignment agency views it. The assignment agency is generally the distributor of a resource, the one responsible for its creation and distribution, much like the publisher is responsible for the assignment of the ISBN. The decision about "sameness" that an assignment agency makes may not be appropriate for all purposes. What algorithm the naming authority uses for determining whether two resources are the same and should have the same URN is purely a decision of that naming authority. Recently the concept of a "Location independent file name (LIFN)" was introduced as yet another standard, which would be used for each valid instance of an object with a description of that object available from the naming authority. Several LIFN's could be associated with one URN.

Until some of these issues are resolved, it does not seem appropriate to find a place in USMARC for the URN. If the URN is defined as location independent, it should be at the record level, i.e. in a separate field, probably in the 0XX block of standard numbers. If each instance of a resource is given a different URN, it may be more desirable to include it in the 856 Electronic Location and Access field. Alternatively, the URL subfield ($u) could be redefined as "URI", to contain either a URL or URI. Since the standard requires a wrapper, "URL:" or "URI:" as part of the data to identify the element, and the subfield is repeatable, both elements could be identified in the same subfield.

7. Data Elements for Online Systems and Services

Following is a list of data elements needed in USMARC records for online systems and services. It is repeated from Discussion Paper No. 69.

Data Element	USMARC Field
Name of the Resource	Bib./CIF 245 (Title)
Acronym/Initialism	Bib. 246 (Varying Form of Title: Format Integration)
Producer	Bib. 260 or 245 $c (Statement of responsibility)
Distributor of the Resource	Bib./Hold. 856 (Electronic Location and Access)
Location	Bib./Hold. 856 $n (Name of host)
Contact Name and Address	856 $m
Network Address(es)	Bib./Hold. 856 $a (Host name)
Hours of Service	CIF 301 $a (Hours, etc.)
Telephone	CIF 270 $k (Telephone number) or $j (Specialized telephone number)
Fax	CIF 270 $l (Fax)
Network Access Instructions	Bib./Hold. 856
Terminal Emulation Supported	Bib./Hold. 856 $t
Logon/Subscription Instructions	Bib./Hold. 856 $l, $z (Logon/login or Public note)
Logoff/Unsubscribe Instructions	Bib./Hold. 856 $z (Public note)
Type of the Resource	Code "i" in 008/26 for computer file (proposed here) Bib. 516 for specific type (e.g. Online Public Access Catalog, Computer Forum, Bulletin Board, etc.)
Size of Resource	Bib. 300 (for number of records, etc.); Bib./Hold. 856 $s (File size if appropriate)
Frequency of Update	Bib. 310 (Current Frequency)
Language of Resource	Bib./CIF 546 (Language Note)
Profile of Resource	Bib./CIF 520 (Summary, Abstract)
Audience	Bib./CIF 521 (Target Audience)
Restrictions on Access	Bib. 506 (Restrictions on Access)
Authorization	Bib. 506 $e (Authorization)
Source Machine	Bib. 538 (Technical Details)
Cost for Use	CIF 531 $b (Fees)
Coverage	Bib. 513 (Type of Report and Period Covered)
Indexing Terms	Bib./CIF 6XX (Subject added entries)
Databases Available	Bib./CIF 505 (Contents)
Other Providers of Database	Bib. 775 (Other Edition Entry)
Documentation Available	Bib. 556 (Information about Documentation Note
Responsibility for Record Maintenance	040 (Cataloging Source)
Date/Time of Last Update of Directory Information	005 (Date and Time of Last of Last Transaction)
Local Access Information and Guidelines	Bib./Hold. 856 $z

7.1 Internet projects for directory services

The Internet Anonymous FTP Archives (IAFA) Working Group of the Internet Engineering Task Force (IETF) has established "cataloging" templates for anonymous FTP archive services which would have a standardized format on the network. The document is being circulated as an RFC (Request for Comment). In addition, the Bunyip Internet Directory Project (BIDP) is running a pilot project to assess the viability of collecting cataloging-type information directly from the Internet. This information would be made available to the Internet community. Its purpose is to identify, describe, characterize, and provide contact information for files and services not available through simple filename searches. Operators of Internet services and anonymous FTP archive site administrators have prepared information for automated gathering of information. The administrators would maintain their own information about their services, and retrieval of the information would be automated. Templates will be automatically retrieved, collated and indexed in a form searchable by users on the network. The general user should be able to query this database and obtain descriptive information about freely available files and services on the network worldwide.

The following example is from a message distributed on Mar. 21, 1994 about the Bunyip project. The corresponding USMARC fields are included (if those proposed in this paper are approved).

A typical "services" template might look like this (no offense to the Census Bureau intended :-)

Data Element	Example	USMARC Field
Template-Type:	SERVICES	008/26 code "i"; 516
Name:	Census Bureau information server	245
Host-Name:	census.ispy.gov	856$a
Host-Port:	1234	856$p
Protocol:	telnet	856 1st indicator or $2
Admin-Name:	Jay Bond	270$p
Admin-Postal:	PO Box. 42, A Street Washington DC, USA 20001	270$a-$e
Admin-Work-Phone:	+1-202-222-3333	270$k
Admin-Work-Fax:	+1 202 444 5555	270$l
Admin-Email:	jb007@census.ispy.gov	270$m
Description:	This server provides information from the latest USA Census Bureau statistics (1990). Type "help" for more information.	520
Authentication:	Once connected type your email address at the "login:" prompt. No password is required.	856$z
Registration:	No formal registration is required	856$z
Charging-Policy:	There is no charge for the use of this service	531
Access-Times:	9:00 EST / 17:00 EST	301
Access-Policy:	This service may not be used by sites in the Republic of the VTTS	506
Keywords:	census, population, 1990, statistics	650; 653
Last-Modified-Name:	Miss Moneypenny	[not clear why needed]
Last-Modified-Email:	m.moneypenny@census.ispy.gov	[not clear why needed]
Last-Modified-Date:	Wed, 1 Jan 1970 12:00:00 GMT	005

8. PROPOSED CHANGES

The following is presented for consideration:

- In the USMARC Bibliographic Format, computer files specifications, add the following code in 008/26 (Type of computer file):
  i Online system or service
- In 008/26 change the description of e (Bibliographic data) to distinguish from i.
  See Attachment A for a description of this field if this proposal is approved.

- In the USMARC Bibliographic Format, define the following Community Information Format fields for use in bibliographic records:
  270 Primary address

  301 Hours, etc. (Option A)
  -or-
  307 Hours, etc. (Option B)

  531 Eligibility, Fees, Procedures Note
  See Attachment B for a list of subfields in these fields.

- In the USMARC Holdings/Bibliographic Formats, define the following subfield in Field 856 (Electronic Location and Access)
  $w Record control number
  See Attachment C for a description of this field if this proposal is approved.

- Attachment D contains some examples of records for online systems and services.

ATTACHMENT A

008/26 Type of computer file
(006/09)

Codes

a	Numeric data	g	Game
b	Computer program	h	Sound
c	Representational	<i	Online system or service>
d	Document	m	Combination
e	Bibliographic data	u	Unknown
f	Font	z	Other

CHARACTER POSITION DEFINITION AND SCOPE

A one-character alphabetic code indicates the type of computer file being described. The specific type of file is also described in textual form in field 516 (Type of Computer File or Data Note).

GUIDELINES FOR APPLYING CONTENT DESIGNATORS

■ CODES

a - Numeric data

Code a indicates a file that contains mostly numbers or representation by numbers, such as records containing all information on student test scores, all information on football team statistics, etc. The information may be original surveys and/or information that has been summarized or statistically manipulated.

008/26 a
516 ##‡aNumeric data

b - Computer program

Code b indicates a file containing an ordered set of instructions directing the computer to perform basic operations and identifying the information and mechanisms required. This category includes videogame and microcomputer software and computer models. Some types of computer programs (e.g., game, font) are identified by separate codes in this character position.

008/26 b
516 ##‡aComputer programs

c - Representational

Code c indicates a file that contains pictorial graphic data that can be manipulated in conjunction with other types of files to produce graphic patterns that can be used to interpret and give meaning to the information. It does not include a document in image format.

008/26 c
516 ##‡aGraphic data (Architectural drawings)

BEST COPY AVAILABLE

d - Document

Code d indicates a file that contains mostly alphabetic information (words or sentences) converted into a coded format that can be processed, sorted, and manipulated by machine, and then retrieved in many optional formats. This category includes such information as records containing full text of documents. It includes language material intended to constitute a textual document, whether represented as ASCII or image data.

008/26 d
516 bb‡aText (Law reports and digests)

e - Bibliographic data

Code e indicates that the item consists of data with bibliographic citations. This includes <data from> library catalogs or citation databases. The data may be in a structured or unstructured form.

008/26 e
516 bb‡a[Library catalog] <Bibliographic records>
Describes a catalog record for a retrospective file of citations from Dissertation abstracts.

f - Font

Code f indicates a file contains information for a computer to produce fonts.

008/26 f
516 bb‡aFonts (Bitmapped and PostScript)

g - Game

Code g indicates that a file is a game, intended for recreational or educational use. Generally games consist of text and software. A videogame is included here.

008/26 g
516 bb‡aComputer game

h - Sounds

Code h indicates that the file consists of data encoding sounds producible by the computer.

<i - Online system or service

Code i indicates that the record is for an online system or service and may contain nonbibliographic information. An online system or service supports system-based user interaction. Examples of these are records for: online library systems, FTP sites, bulletin boards, network information centers.

008/26 i
516 bb‡aCampus-wide information System>

m - Combination

Code m is used when the item is a combination of two or more of the above types of files.

008/26 m
516 bb‡aComputer programs and text files

u - Unknown

Code u indicates that the type of file is unknown.

ATTACHMENT B

Community Information Fields adapted for use in bibliographic records

Note: [ ] shows deletion from CIF field; < > shows addition from CIF field

270 Primary Address (R)

Indicators

First Undefined
b Undefined

Second Undefined
b Undefined

Subfield Codes

ǂa Address (R)
ǂb City (NR)
ǂc State or province (NR)
ǂd Country (NR)
ǂe Postal code (NR)
ǂf Title preceding attention name (NR)
ǂg Attention name (NR)
ǂh Title following attention name (NR)
ǂi Type of address (NR)
ǂj Specialized telephone number (R)
ǂk Telephone number (R)
ǂl Fax number (R)
ǂm Electronic mail address (R)
ǂn TDD or TTY number (R)
ǂp Contact [person](R)
ǂq Title of contact [person] (R)
<ǂ4 Relator code (R)>
ǂ6 Linkage (NR)

FIELD DEFINITION AND SCOPE

This field contains the address (as well as electronic access data, such as telephone, fax, TTY, etc. numbers) associated with the item described in the record. The field may be repeated fro additional addresses, e.g., when different contacts are associated with the item. The relationship with the item is indicated by a code in subfield ǂ4.

301 (or 307) Hours, Etc. (R)

Indicators

First Display constant controller
b No information provided
8 No display constant generated

Second Undefined
b Undefined

Subfield Codes

ǂa Hours (NR)
ǂb Additional information (NR)
<ǂc Time Differential Factor (NR)>
ǂ6 Linkage (NR)

FIELD DEFINITION AND SCOPE

This field contains information as to the days and/or times pertaining to the operations of the entity described in the record (such as when hours of availability of the service), as well as to special information, e.g., the size of the staff.

When displayed/printed as a note, hours, etc. information is in some instances preceded by an introductory term or phrase that is generated based on the first indicator value.

531 Eligibility, Fees, Procedures Note (R)

Indicators

First Undefined
b Undefined

Second Undefined
b Undefined

Subfield Codes

‡a Eligibility (NR)
‡b Fee (R)
‡c Admission procedures (NR)
‡d Documents required (NR)
‡e Waiting list (NR)
‡f Waiting period (NR)
‡6 Linkage (NR)

FIELD DEFINITION AND SCOPE

This field contains certain information regarding the community information entity, e.g., who is eligible to use the services, attend the event, join the organization; the fee involved; admission procedures; the waiting period, etc.

ATTACHMENT C

856 Electronic Location and Access (R)

Indicators

First	Access method
0	Email
1	FTP
2	Remote login (Telnet)
7	<Other access method> [Source] specified in subfield ǂ2

Second	Undefined
b	Undefined

Subfield Codes

ǂa	Host name (R)	ǂn	Name of location of host in subfield ǂa (NR)
ǂb	IP address (NR)	ǂo	Operating system (NR)
ǂc	Compression information (R)	ǂp	Port (NR)
ǂd	Path (R)	ǂq	File transfer mode (NR)
ǂf	Electronic name (R)	ǂs	File size (R)
ǂg	Electronic name–End of range (R)	ǂt	Terminal emulation (R)
ǂh	Processor of request (NR)	ǂu	Uniform Resource Locator (R)
ǂi	Instruction (R)	<ǂw	Record control number (R)>
ǂk	Password (NR)	ǂx	Nonpublic note (R)
ǂl	Logon/login (NR)	ǂz	Public note (R)
ǂm	Contact for access assistance (R)	ǂ2	[Source of] Access <method> (NR)
		ǂ3	Materials specified (NR)

FIELD DEFINITION AND SCOPE

This field contains the information required to locate an electronic item. The information identifies the electronic location containing the item or from which it is available. It also contains information to retrieve the item by the access method identified in the first indicator position. The information contained in this field is sufficient to allow for the electronic transfer of a file, subscription to an electronic journal, or logon to a library catalog. In some cases, only unique data elements are recorded which allow the user to access a locator table on a remote host containing the remaining information needed to access the item.

Field 856 is repeated when the location data elements vary (subfields ǂa, ǂb, ǂd) and when more than one access method may be used. It is also repeated whenever the electronic filename varies (subfield ǂf), except for the situation when a single intellectual item is divided into different parts for online storage or retrieval.

856 0b ǂakeptvm.bitnet ǂfacadlist file1 ǂs34,989 bytes ǂfacadlist file2 ǂs32,876 bytes ǂfacadlist file3 ǂs23987 bytes

ǂt - Terminal emulation

Subfield ǂt contains the terminal emulation supported when necessary to specify for remote login (first indicator contains value 2 (Remote login (Telnet)).

856 2b ǂamaine.maine.edu ǂnUniversity of Maine ǂt3270

ǂu - Uniform Resource Locator

Subfield ǂu contains the Uniform Resource Locator (URL), which provides standard syntax for locating an object using existing Internet protocols. Field 856 is structured to create a URL from separate subfields. Subfield ǂu may be used instead of those separate subfields or in addition to them. It might be desirable to include subfield ǂu and the other subfields if a user display is desired as well as a URL. The field is repeated if more than one URL needs to be recorded.

856 1b ǂuURL: ftp://path.net/pub/docs/urn2urc.ps

< ǂw - Record control number

Subfield ǂw contains the system control number of the related record preceded by the USMARC code, enclosed in parentheses, for the agency to which the control number applies. (The source of this code is *Symbols of American Libraries* that is maintained by the Library of Congress.)>

ǂx - Nonpublic note

Subfield ǂx contains a note relating to the electronic location of the source identified in the field. The note is written in a form that is not adequate for public display or contains processing information about the file at the location specified.

856 1b ǂawuarchive.wustl.edu ǂcdecompress with PKUNZIP.exe ǂd/mirrors2/win3/games ǂfatmoids.zip ǂxcannot verify because of transfer difficulty

ǂz - Public note

Subfield ǂz contains a note relating to the electronic location of the source identified in the field. The note is written in a form that is adequate for public display.

ǂ2 - [Source of] Access method

Subfield ǂ2 contains the [source of] access <method> when the first indicator position contains value 7 (Access method specified in subfield ǂ2). This subfield may include access methods other than the three main TCP/IP protocols specified in the first indicator. This subfield is controlled by an authoritative list maintained at the Library of Congress.

ǂ3 - Materials specified

Subfield ǂ3 contains information that specifies the part of the [bibliographic item]<electronic resource> to which the field applies.

ATTACHMENT D

EXAMPLE 1: InterNIC Information Services

008 yymmddnnnnn####caun########i#########
245 $a InterNIC Information Services Info Source $c InterNIC Information Services
246 $b InterNIC
260 $a San Diego, Calif. : $b InterNIC Information Services
270 $k 800.444.4345 $k 619.455.4600 $l 619.455.3900
310 Biweekly
500 Title from the opening menu
516 Network Information Center
520 Info Source is a collection of information about Internet and all the services offered by InterNIC Information Services.
521 Network users
538 Access through computer network
856 0 $a is.internic.net $h mailserv $i send help
856 0 $a is.internic.net $h mailserv $i Index
856 1 $a is.internic.net $k your email address $l anonymous
856 2 $a is.internic.net $l gopher $t vt100
856 7 $a is.internic.net $2 WAIS client $2 internic-infosource
856 7 $a is.internic.net $2 Gopher client

EXAMPLE 2: CARL

008 yymmddnnnnn####coun########\########
245 $a CARL
246 $a Colorado Alliance of Research Libraries
260 $a Denver, CO $bCARL Systems Inc.
270 $k303.758.3030 $l303.785.0606
301 $a 24 hours
310 $a Updated daily
500 $a 40 subscribers, 1500 dedicated terminals, 5,948,644 records
505 $a ERIC, Choice, UnCover, CONSER, Magazine Index, Business Index
506 Public access to catalog
506 $eNo password required for public access
513 Holdings vary according to individual library owners
516 $a Online Public Access Catalog
520 $a Public access catalog covering holdings of most academic, public, and special libraries in Colorado as well as other libraries throughout the United States. Also provides access to several specialized databases
521 $a Students, researchers, faculty, public
531 $a No cost to users except for UnCover document delivery service and certain specialized databases funded by individual library owners
546 $a English
556 $a Documentation available for Circulation, Database Maintenance, Public Access Catalog (PAC), Reserve and Serials systems
653 $a Library catalogs
653 $a Online catalogs
653 $a Citation indexes
653 $a Data bases
856 2 $a pac.carl.org $b 192.54.81.128 $m CARL Situation Room $m help@CARL.org $nCARL Systems Inc., Denver, CO

EXAMPLE 3: LIBRARY OF CONGRESS INFORMATION SYSTEM

008 940505m19689999dcudr\\g\\\i\f\\\\\\eng\\
040 $a DLC$c DLC
042 $a gils
245 10$a LOCIS, Library of Congress Information System.
246 13$a LOCIS
246 13$a Library of Congress Information System
246 13$a LC online search
260 $a Washington, DC :$b Library of Congress,$c 1968-
300 $a records <26+ million>
301 $a M-F 6:30 AM to 9:30 PM, Sa 8:00 AM to 5:00 PM Su 1:00 PM to 5:00 PM; closed on national holidays $c -0400
310 $a Updated daily
520 $a A conglomeration of files containing more than 26 million records, the earliest of which was created in 1968.
531 $a No cost to users.
610 20$a Library of Congress$xInformation services.
710 2 $a Library of Congress.
856 2 $a locis.loc.gov $b 140.147.254.3 $m lconline@seq1.loc.gov $nLibrary of Congress, Washington, DC, 20540

Attachment P

BUILDING A POLICY FOR INFORMATION TECHNOLOGY STANDARDS

A WORKING PAPER

William E. Moen
<wemoen@mailbox.syr.edu>
School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244

Introduction

A previous paper (Moen, 1994) argued that there is a link between information technology standards and broader information policy goals. Recent policy initiatives (e.g., the revised Office of Management and Budget [OMB] Circular A-130, information infrastructure legislation, the Government Information Locator Service [GILS]) indicate an increasing recognition of the role of standards in the use and flow of information. Thus, this may be the appropriate time to call for an information technology standards policy for the Federal government. Such a policy may help move the Federal government towards the interconnection and interoperable horizon of the evolving National Information Infrastructure (NII). Preliminary to the development of a standards policy, however, three activities must be undertaken as a means of advancing such a policy:

- Determining a policy goal that guides and directs standards activities
- Adopting a framework for selecting appropriate standards
- Developing a management strategy that assists agencies and the Federal government in its standards activities and use.

The following sections outline the substance of these activities and offer preliminary recommendations.

Policy Goal of Information Technology Standardization

There is a need for an overarching policy goal for information technology standards. Revisions to A-130 (Office of Management and Budget, 1994) direct agencies to deploy information technology and build information systems in the context of or on a migration path to an open systems environment. The emerging NII is assumed to be built as an open environment, connecting a wide array of information appliances across a variety of network and telecommunications paths. The recent report on Federal networking requirements discussed the need for a goal, if not a framework, to guide agencies in their choice of standards-based technologies and suggested the general goal of interoperabililty (Federal Internetworking Requirements Panel (1994).

For the Federal government, an appropriate overarching policy goal for information technology standards is the achievement of an open systems environment. An OMB document describes what open systems can offer users and how it can benefit them (Office of Management and Budget. Office of Information and Regulatory Affairs, 1992, pp. 23-24). Open systems enable users to achieve the following:

- Portability: The ability to use, or migrate, systems software, applications software, and data across different computing platforms from multiple vendors

- Interoperability: The ability to have applications and computers from different vendors work together on a network
- Scalability: The ability to use the same applications and systems software on all classes of computers from desktop workstations to supercomputers
- Common Programming Interfaces: The ability to develop applications based on a set of standards programming tools which can be easily transferred across platforms
- Common User Interfaces: The ability to create applications with a similar "look and feel" so that users can easily learn new applications after understanding the first.

The concept of open systems can be seen as an evolution of the goals expressed by the Brooks Act (P.L. 89-306) since the potential benefits that result from selecting information technology that conform to established open systems architecture and standards include: vendor independence; protection of investment; fast time to market; improved systems integration; and lower costs. In addition, many agency information resources managers recognize the utility of open systems and standards-based solutions to information technology issues; a survey of information resources managers found 70% support for National Institute of Standards and Technology (NIST) open systems initiatives (Information Technology Association of America, 1992).

A policy goal of building open systems would challenge agencies to deploy technology that is portable, interoperable, scalable, etc. By focusing the goal on performance requirements (i.e., create an open systems architecture) rather than design or technology-specific requirements (e.g., use specific national and international standards) agencies will have a flexibility to respond to the fast pace of technology change. The agencies will be responsible for satisfying the performance requirements by, for example, deploying information systems that are demonstrably interoperable within the agency, across the Federal government, and within the larger national and global information environment.

Framework for Selecting and Deploying Standards

The second component for developing a policy addresses the need for guidance to the government and individual agencies in selecting appropriate standards. An overarching policy goal of creating open systems provides a baseline for standards choice. If information technology standards, however, are to assist in achieving broader information policy goals, more specific guidance on selecting information handling standards is required. Past and current policy instruments that address the use of information technology or other information handling standards, however, have yet to envision such uses of standards.

Circular A-130 placed increased emphasis on the concept of information life cycle in managing information resources. The information life cycle can be used as a framework for identifying and selecting appropriate standards (Spring & Bearman, 1988). Information life cycle refers to the stages through which information passes (Office of Management and Budget, 1994). These include:

- Creation or collection
- Processing
- Dissemination
- Use
- Storage
- Disposition.

While certain of these stages are logically and temporally prior to others (e.g., information must be collected or created before it can be disposed), the nature of electronic information exposes the nonlinear characteristic of the life cycle (Hernon, 1994). Electronic information that is available from one agency may be recollected and reprocessed by another agency or organization before being disseminated.

Figure 1 presents examples of standards (both technology and others) involved in the information life cycle stages. This is not an exhaustive list of activities within each stage or relevant standards but is intended to show that processes involved in each stage are subject to standardized solutions.

Figure 1
Relevant Standards for Information Life Cycle Stages

Life Cycle Stage	Information Activities	Relevant Standards
Creation or Collection	Document Preparation	Standard Generalized Markup Language (SGML)
Processing	Organizing	Anglo-American Cataloguing Rules; Indexes
Dissemination	Communications, Networking	USMARC; OSI, TCP/IP
Use	Information Retrieval	Z39.50; SQL
Storage	Optical disks	CD-ROM
Disposition	Microfilming	Preservation standards

The intersection of the information life cycle perspective and technical standards allows one to begin to think how value can be added through the use of accepted standards. As one example, if the Standard Generalized Markup Language (SGML), which provides a way of tagging the data in and identifying the structure of a document, is used at the time of creation rather than using proprietary wordprocessing software or creating ASCII text files, the resulting SGML document can be output in a variety of ways and SGML enables more robust manipulation and searching of the document at later stages in the life cycle.

For each stage of the life cycle, and especially when dealing with electronic information, standards play an essential role in shaping how they are done. One might say that the technology shapes how they are done, but since standards are deeply embedded in the information technology, it may be appropriate to focus more on the standards and less on the technology.

A Management Strategy for Information Technology Standardization

A overarching policy goal and a framework for identifying appropriate standards or the need for specific standards must be accompanied by a management and organizational response to standardization. Developing a management strategy for information technology standards is the third activity that lays the groundwork for a policy on standards.

Recent writers have discussed the need for organizations to establish a mechanism to carry out standards oversight and coordination (Ritterbusch, 1990; Betancourt, 1993). There is no clear indication that Federal agencies have implemented such procedures. Whether a specific "standards program" or "standards organization" is established, a number of functions and activities need to be considered. Figure 2 presents these functions and activities as outlined by Ritterbusch (1990).

Figure 2
Aspects of a Standards Program

- Identify needs
  - monitor external motivators for standards
  - keep pace with technology
- Take appropriate action
  - adopt or adapt existing standards
  - prepare new standards
  - assure technical validity
- Distribute and maintain standards
  - updating
  - maintenance
- Implement standards
  - maximize usage
  - mandatory usage
  - ongoing implementation
- Other functions
  - train standards users
  - provide advisory services
  - external liaison
  - monitor activities of external standards organizations.

Information resources management (IRM) units are appropriately placed, as well as directed by the Paperwork Reduction Act and its Reauthorization (P.L. 96-511, P.L. 99-500) and A-130, to lead agency planning for standardization related to information technology, information services, and the electronic delivery of services. Ideally, there should be a close connection between IRM responsibilities for diffusion of innovation related to information technology and standards awareness. Innovative standards-based solutions can provide long term efficiencies and increased effectiveness. This requires, however, that IRM staff monitor the changing technology environment and carry out the functions outlined in Figure 2 as part of a consciously designed management strategy for standardization. A planning process for standardization will be an essential component of the strategic planning for information technology called for in the 1994 revisions to A-130.

Although individual agencies must respond to the issues of standardization, there is also a need for government-wide response. A government-wide management strategy will need to address: mechanisms (e.g., interagency groups) for sharing information about agency standards activities and for coordination among those who are actually deploying the technology; guidance and oversight of agency standards programs (e.g., OMB budget review process); coordination of standards implementation across Federal agencies (e.g., a single agency such as NIST or OMB or an interagency group such as the Interagency Committee on Standards Policy); roles and responsibilities for specific agencies in standards education and training; a coherent and effective way to measure, evaluate, and achieve agency compliance with and use of standards. A workable management strategy must include a combination of rewards and incentives, as well as acknowledging agency self-interest and the need for tangible benefits.

Finally, a realistic management strategy for standardization will acknowledge the time-frame appropriate to standards. While there may be short term benefits, it is in the longer term that benefits accrue; as the director of the National Information Standards Organization pointed out, "standards are a long-term investment — both for development and implementation....We must not lose sight of the long term benefits of standards, and we must make the long term commitment to realize those benefits" (Harris, 1994, p. 32) .

The three activities tentatively proposed here will need further elaboration. Additional research on other key policy areas (e.g., which information technology standards to choose; how to coordinate the use of information technology standards; and who should enforce compliance with information technology standards; see Moen, [1994] for a description of these policy areas) will provide a point of departure for articulating the substance of these activities and lay the foundation for developing policies for information technology standardization.

For nearly thirty years, the Federal government has expressed its commitment to the use of standards in the procurement, adoption, implementation, and use of information technology. While cost efficiency in the use of information systems can still be an important objective, the use of these systems now has become an accepted, albeit critical, component of information handling in the Federal government. New goals and objectives are needed to guide the adoption and use of information technology standards. Data sharing and interoperability have been important considerations in the past, but the development of national electronic networks, the electronic delivery of government information and services, and the commitment of the current administration to the NII makes this an especially opportune time to focus on information technology standards as the linchpin to successful information infrastructures, whether agency, government-wide, national or international infrastructures. Linking the use of standards to the achievement of long-standing information policy goals provides the basis on which to develop a coherent policy for Federal information technology standards.

REFERENCES

Betancourt, Diego. (1993, November 19). Strategic standardization management: A strategic macroprocess approach to the new paradigm in the competitive business use of standardization. [Available from the author: Diego Betancourt, Manager, Office of Strategic Standardization, Polaroid Corporation, 656 Technology Square, Cambridge, MA 02139].

Federal Internetworking Requirements Panel. (1994, May 31). Report of the federal internetworking requirements panel: Prepared for the National Institute of Standards and Technology. Gaithersburg, MD: National Institute of Standards and Technology. Available via anonymous FTP at <osi.ncsl.nist.gov> in directory /pub/firp/

Harris, Patricia. (1994). National standards development in the United States: Private sector interests in the public sector environment. In William E. Moen (Ed.), Scholarly Information and Standardization (pp. 21-32). Bethesda, MD: NISO Press.

Hernon, Peter. (1994). Information life cycle: Its place in the management of U.S. government information resource. Government Information Quarterly 11 (2): 143-170.

Information Technology Association of America. (1992, July). Key issues in federal information technology: Third annual ITAA survey of senior information resources management officials. Arlington, VA: Information Technology Association of America.

Moen, William E. (1994). Information technology standards: A component of federal information policy. Government Information Quarterly 11 (4): 357-371.

Office of Management and Budget. (1994, July 25). Management of federal information resources [Circular no. A-130]. Federal Register 59: 37906-37928.

Office of Management and Budget. Office of Information and Regulatory Affairs. (1992, March). Budget examining techniques for evaluating information technology investments. Washington,DC: Office of Management and Budget.

Ritterbusch, Gerald H. (1990). Functions of a standards program. In Robert B. Toth (Ed.), Standards Management: A Handbook for Profits (pp. 19-34). New York: American National Standards Institute.

Spring, Michael B. & Bearman, Toni Carbo. (1988, Fall). Information standards: Models for future development. Book Research Quarterly 4: 38-47

Attachment Q

THE GILS FORUM: AN ELECTRONIC DISCUSSION GROUP

A GILS Forum has been established to provide a forum for discussions related to the GILS initiative. This electronic forum is hosted by the Coalition for Networked Information (CNI). The following provides some basic information about the scope of the GILS Forum as well as directions for subscribing to it.

About the GILS Forum

GILS is a public, moderated computer forum open to anyone interested in discussing development, implementation, deployment, policy, and technical issues surrounding the Government Information Locator Service (GILS) initiative. The GILS forum provides the opportunity for United States and other government and non-government organizations' staff, state and local government agencies, Z39.50 implementors, librarians, information service providers, public interest groups, the general public, and others to participate in discussions related to the GILS. The GILS forum will also announce and provide electronic access to documents related to the GILS initiative.

Since the GILS architecture is an application to identify and locate information resources in a variety of networked information environments , the broad scope of the discussion on the GILS forum may include topics such as the use of GILS in an international context, the technical implementation details of GILS, the creation of GILS records, development of generalized Uniform Resource Identifiers, and the use of ANSI/NISO Z39.50 in the GILS. In addition, the GILS forum provides a point of contact for vendors who are developing GILS products and services and potential users of these products and services.

To Subscribe to the GIS Forum

To join the Government Information Locator Service electronic forum, please send a SUBSCRIBE command (as an e-mail note) to the address:

LISTPROC@CNI.ORG

subscribe <listname> <your real name>

e.g. SUBSCRIBE GILS John Doe

To Send Mail to the GILS Forum

To participate in the list discussions, please send your mail to:

GILS@CNI.ORG

To Leave the GILS Forum

If you are a subscriber of the GILS forum, and you wish to leave the forum, please send either an UNSUBSCRIBE or a SIGNOFF command (as an e-mail note) to the address LISTPROC@CNI.ORG

unsubscribe <listname> or signoff <listname>

e.g. UNSUBSCRIBE GILS SIGNOFF GILS

About UNIX-LISTPROCESSOR

This forum is operated in the unix environment, but is not a simple mail-reflector. The Unix-Listprocessor (formerly called Unix-listserv), programmed by Anastasios C. Kotsikonas (Copyright (c) 1991,1992,1993,1994), has many features similar to the L-Soft (formerly Revised [BITNET] LISTSERV system, although the syntax of some commands may vary slightly. To begin learning more about Unix-Listprocessor, send one or more of the following commands to LISTPROC@CNI.ORG

HELP
GET LISTPROC REFCARD
LISTS

No subject line is required (or suggested) when communicating with the Unix-Listproc (LISTPROC@CNI.ORG), although if you include one it will be ignored. After your request has been processed, you will receive a message from the Unix-Listproc giving brief information regarding the commands this system supports.

For Further Information on this Forum

All questions regarding the substance of, or policies related to, the discussions on this forum should be sent to:

William E. Moen
Research Associate and GILS Forum Moderator
School of Information Studies
4-206 Center for Science and Technology
Syracuse University
Syracuse, NY 13244
(315) 445-0015/443-4508
Internet: wemoen@mailbox.syr.edu

and queries regarding difficulties with mail or requests for technical assistance should be sent to the Coalition's Systems Coordinator:

Craig A. Summerhill
Systems Coordinator and Program Officer
Coalition for Networked Information
21 Dupont Circle, N.W.
Washington, D.C. 20036
(202) 296-5098
Internet: craig@cni.org

July 27, 1994